

B3

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
15 August 2002 (15.08.2002)

PCT

(10) International Publication Number  
**WO 02/063005 A2**

(51) International Patent Classification<sup>7</sup>: **C12N 15/12**,  
C07K 14/47, A01K 67/00, C12Q 1/68, A61K 38/17,  
G01N 33/50, C07K 16/18

(21) International Application Number: PCT/US02/03813

(22) International Filing Date: 6 February 2002 (06.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/266,910	6 February 2001 (06.02.2001)	US
60/276,891	16 March 2001 (16.03.2001)	US
60/276,855	16 March 2001 (16.03.2001)	US
60/279,760	28 March 2001 (28.03.2001)	US
60/283,818	13 April 2001 (13.04.2001)	US
60/285,405	20 April 2001 (20.04.2001)	US

(71) Applicant (for all designated States except US): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **DAS, Debopriya** [IN/US]; 1267 Parkington Avenue, Sunnyvale, CA 94087 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **ARVIZU, Chandra** [US/US]; 1706 Morocco Drive, San Jose, CA 95125 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **LU, Yan** [US/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **DING, Li** [CN/US]; 3353 Alma Street, #146, Palo Alto, CA 94306 (US). **TOWNLEY, David, J.** [GB/GB]; 10 Pentlands Court, Cambridge CB4 1JN (GB). **ELLIOTT, Vicki, S.** [US/US]; 3770 Pollan Place Way, San Jose,

CA 95121 (US). **FORSYTHE, Ian** [US/US]; 308 Roble Avenue, Redwood City, CA 94061 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US). **WARREN, Bridget, A.** [US/US]; 10130 Parkwood Drive #2, Cupertino, CA 95014 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue #71, Palo Alto, CA 94303 (US). **HONCHELL, Cynthia, D.** [US/US]; 400 Laurel Street #203, San Carlos, CA 94070 (US).

(74) Agents: **HAMLET-COX, Diana et al.**; Incyte Genomics, 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/063005 A2

(54) Title: LIPID-ASSOCIATED MOLECULES

(57) Abstract: The invention provides human lipid-associated molecules (LIPAM) and polynucleotides which identify and encode LIPAM. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of LIPAM.

## LIPID-ASSOCIATED MOLECULES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of lipid-associated molecules and to the use of these sequences in the diagnosis, treatment, and prevention of cancers, neurological, autoimmune/inflammatory, gastrointestinal, and cardiovascular disorders, and disorders of lipid metabolism, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of lipid-associated molecules.

### BACKGROUND OF THE INVENTION

Lipids are water-insoluble, oily or greasy substances that are soluble in nonpolar solvents such as chloroform or ether. Neutral fats (triacylglycerols) serve as major fuels and energy stores. Fatty acids are long-chain organic acids with a single carboxyl group and a long non-polar hydrocarbon tail. Long-chain fatty acids are essential components of glycolipids, phospholipids, and cholesterol, which are building blocks for biological membranes, and of triglycerides, which are biological fuel molecules. Lipids, such as phospholipids, sphingolipids, glycolipids, and cholesterol, are key structural components of cell membranes. Lipids and proteins are associated in a variety of ways. Glycolipids form vesicles that carry proteins within cells and cell membranes. Interactions between lipids and proteins function in targeting proteins and glycolipids involved in a variety of processes, such as cell signaling and cell proliferation, to specific membrane and intracellular locations. Various proteins are associated with the biosynthesis, transport, and uptake of lipids. In addition, key proteins involved in signal transduction and protein targeting have lipid-derived groups added to them post-translationally (Stryer, L. (1995) Biochemistry, W.H. Freeman and Co., New York NY, pp. 264-267, 934; Lehninger, A. (1982) Principles of Biochemistry, Worth Publishers, Inc. New York NY; and ExPASy "Biochemical Pathways" index of Boehringer Mannheim World Wide Web site, "http://www.expasy.ch/cgi-bin/search-biochem-index".)

#### Phospholipids

A major class of phospholipids are the phosphoglycerides, which are composed of a glycerol backbone, two fatty acid chains, and a phosphorylated alcohol. Phosphoglycerides are components of cell membranes. Principal phosphoglycerides are phosphatidyl choline, phosphatidyl ethanolamine, phosphatidyl serine, phosphatidyl inositol, and diphosphatidyl glycerol. Many enzymes involved in phosphoglyceride synthesis are associated with membranes (Meyers, R.A. (1995) Molecular Biology and Biotechnology, VCH Publishers Inc., New York NY, pp. 494-501). Phosphatidate is converted to

CDP-diacylglycerol by the enzyme phosphatidate cytidyltransferase (ExPASy ENZYME EC 2.7.7.41). Transfer of the diacylglycerol group from CDP-diacylglycerol to serine to yield phosphatidyl serine, or to inositol to yield phosphatidyl inositol, is catalyzed by the enzymes CDP-diacylglycerol-serine O-phosphatidyltransferase and CDP-diacylglycerol-inositol 3-phosphatidyltransferase, respectively (ExPASy ENZYME EC 2.7.8.8; ExPASy ENZYME EC 2.7.8.11). The enzyme phosphatidyl serine decarboxylase catalyzes the conversion of phosphatidyl serine to phosphatidyl ethanolamine, using a pyruvate cofactor (Voelker, D.R. (1997) *Biochim. Biophys. Acta* 1348:236-244). Phosphatidyl choline is formed using diet-derived choline by the reaction of CDP-choline with 1,2-diacylglycerol, catalyzed by diacylglycerol cholinephosphotransferase (ExPASy ENZYME 2.7.8.2).

Other phosphoglycerides have been shown to be involved in the vesicle trafficking process. Phosphatidylinositol transfer protein (PITP) is a ubiquitous cytosolic protein, thought to be involved in transport of phospholipids from their site of synthesis in the endoplasmic reticulum and Golgi to other cell membranes. More recently, PITP has been shown to be an essential component of the polyphosphoinositide synthesis machinery and is hence required for proper signaling by epidermal growth factor and f-Met-Leu-Phe, as well as for exocytosis. The role of PITP in polyphosphoinositide synthesis may also explain its involvement in intracellular vesicular traffic (Liscovitch, M. et al. (1995) *Cell* 81:659-662).

The copines are phospholipid-binding proteins believed to function in membrane trafficking. Copines promote lipid vesicle aggregation. They contain a C2 domain associated with membrane activity and an annexin-type domain that mediates interactions between integral and extracellular proteins and is associated with calcium binding and regulation (Creutz, C.E. (1998) *J. Biol. Chem.* 273:1393-1402). Other C2-containing proteins include the synaptotagmins, a family of proteins involved in vesicular trafficking. Synaptotagmin concentrations in cerebrospinal fluid have been found to be reduced in early-onset Alzheimer's disease (Gottfries, C.G. et al. (1998) *J. Neural Transm.* 105:773-786).

The phosphatidylinositol-transfer protein Sec14, which catalyses exchange of phosphatidylinositol and phosphatidylcholine between membrane bilayers *in vitro*, is essential for vesicle budding from the Golgi complex. Sec14 includes a carboxy-terminal domain that forms a hydrophobic pocket which represents the phospholipid-binding domain (Sha, B. et al. (1998) *Nature* 391:506-510). Sec14 is a member of the cellular retinaldehyde-binding protein (CRAL)/Triple function domain (TRIO) family (InterPro Entry IPR001251, <http://www.ebi.ac.uk/interpro>).

#### Sphingolipids

Sphingolipids are an important class of membrane lipids that contain sphingosine, a long chain amino alcohol. They are composed of one long-chain fatty acid, one polar head alcohol, and sphingosine or sphingosine derivatives. The three classes of sphingolipids are sphingomyelins, cerebroside, and gangliosides. Sphingomyelins, which contain phosphocholine or phosphoethanolamine as their head group, are abundant in the myelin sheath surrounding nerve cells. Galactocerebroside, which contains a glucose or galactose head group, are characteristic of the brain. Other cerebroside are found in non-neural tissues. Gangliosides, whose head groups contain multiple sugar units, are abundant in the brain, but are also found in non-neural tissues.

#### Glycolipids

Glycolipids are also important components of the plasma membranes of animal cells. The most simple glycolipid is cerebroside which comprises only a single glucose or galactose sugar residue in addition to the lipid component. Gangliosides are glycosphingolipid plasma membrane components that are abundant in the nervous systems of vertebrates. Gangliosides are the most complex glycolipids and comprise ceramide (acylated sphingosine) attached to an oligosaccharide moiety containing at least one acidic sugar residue (sialic acid), namely *N*-acetylneuraminate or *N*-glycolylneuraminate. The sugar residues are added sequentially to ceramide via UDP-glucose, UDP-galactose, *N*-acetylgalactosamine, and CMP-*N*-acetylneuraminate donors. Over 15 gangliosides have been identified with  $G_{M1}$  and  $G_{M2}$  being the best characterized (Stryer, L. (1988) Biochemistry, W.H. Freeman and Co., Inc., New York, pp. 552-554).

Gangliosides are thought to play important roles in cell surface interactions, cell differentiation, neurite outgrowth, the triggering and modulation of transmembrane signaling, mediating synaptic function, neural repair, neurite outgrowth, and neuronal death (Hasegawa, T. et al. (2000) *J. Biol. Chem.* 275:8007-8015). While the presence of gangliosides in the plasma membrane is important for orchestrating these events, the subsequent removal of carbohydrate groups (desialylation) by sialidases also appears to be important for regulating neuronal differentiation.

Specific soluble N-ethylmaleimide-sensitive factor attachment protein (SNAP) receptor (SNARE) proteins are required for different membrane transport steps. The SNARE protein Vti1a has been colocalized with Golgi markers while Vti1b has been colocalized with Golgi and the trans-Golgi network of endosomal markers in fibroblast cell lines. A brain-specific splice variant of Vti1a is enriched in small synaptic vesicles and clathrin-coated vesicles isolated from nerve terminals. Vti1a-beta and synaptobrevin are integral parts of synaptic vesicles throughout their life cycle. Vti1a-beta functions in a SNARE complex during recycling or biogenesis of synaptic vesicles (Antonin, W. et al. (2000) *J. Neurosci.* 20:5724-5732).

Sialidases catalyze the first step in glycosphingolipid degradation, removing carbohydrate moieties from gangliosides. These enzymes are present in the cytosol, lysosomal matrix, lysosomal membrane, and plasma membrane (Hasegawa, T. et al. (2000) *J. Biol. Chem.* 275:8007-8015).

Hallmark features of sialidases include a transmembrane domain, an Arg-Ile-Pro domain, and three  
5 Asp-box sequences (Wada, T. (1999) *Biochem. Biophys. Res. Commun.* 261:21-27).

During normal neuronal development, pyramidal neurons of the cerebral cortex participate in a single burst of dendritic sprouting immediately following nerve cell migration to the cortical mantle. Cells undergoing dendritogenesis are characterized by increased expression of G<sub>M2</sub> ganglioside which decreases following dendritic maturation. Evidence suggests that no new primary dendrites are  
10 initiated following the initial burst.

### Cholesterol

Cholesterol, composed of four fused hydrocarbon rings with an alcohol at one end, moderates the fluidity of membranes in which it is incorporated. In addition, cholesterol is used in the synthesis of steroid hormones such as cortisol, progesterone, estrogen, and testosterone. Bile salts derived from  
15 cholesterol facilitate the digestion of lipids. Cholesterol in the skin forms a barrier that prevents excess water evaporation from the body. Farnesyl and geranylgeranyl groups, which are derived from cholesterol biosynthesis intermediates, are post-translationally added to signal transduction proteins such as Ras and protein-targeting proteins such as Rab. These modifications are important for the activities of these proteins (Guyton, A.C. (1991) *Textbook of Medical Physiology*, W.B. Saunders  
20 Company, Philadelphia PA, pp. 760-763; Stryer, *supra*, pp. 279-280, 691-702, 934).

Mammals obtain cholesterol derived from both *de novo* biosynthesis and the diet. The liver is the major site of cholesterol biosynthesis in mammals. Biosynthesis is accomplished via a series of enzymatic steps known as the mevalonate pathway. The rate-limiting step is the conversion of hydroxymethylglutaryl-Coenzyme A (HMG-CoA) to mevalonate by HMG-CoA reductase. The drug  
25 lovastatin, a potent inhibitor of HMG-CoA reductase, is given to patients to reduce their serum cholesterol levels. Cholesterol derived from *de novo* biosynthesis or from the diet is transported in the body fluids in the form of lipoprotein particles. These particles also transport triacylglycerols. The particles consist of a core of hydrophobic lipids surrounded by a shell of polar lipids and apolipoproteins. The protein components serve in the solubilization of hydrophobic lipids and also  
30 contain cell-targeting signals. Lipoproteins include chylomicrons, chylomicron remnants, very-low-density lipoproteins (VLDL), intermediate-density lipoproteins (IDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL) (Meyers, *supra*; Stryer, *supra*, pp. 691-702). There is a strong inverse correlation between the levels of plasma HDL and risk of premature coronary heart disease.

ApoL is an HDL apolipoprotein expressed in the pancreas (Duchateau, P.N. et al. (1997) J. Biol. Chem. 272:25576-25582).

Most cells outside the liver and intestine take up cholesterol from the blood rather than synthesize it themselves. Cell surface LDL receptors bind LDL particles which are then internalized  
5 by endocytosis (Meyers, supra). Absence of the LDL receptor, the cause of the disease familial hypercholesterolemia, leads to increased plasma cholesterol levels and ultimately to atherosclerosis (Stryer, supra, pp. 691-702).

Proteins involved in cholesterol uptake and biosynthesis are tightly regulated in response to cellular cholesterol levels. The sterol regulatory element binding protein (SREBP) is a sterol-  
10 responsive transcription factor. Under normal cholesterol conditions, SREBP resides in the endoplasmic reticulum membrane. When cholesterol levels are low, a regulated cleavage of SREBP occurs which releases the extracellular domain of the protein. This cleaved domain is then transported to the nucleus where it activates the transcription of the LDL receptor gene, and genes encoding enzymes of cholesterol synthesis, by binding the sterol regulatory element (SRE) upstream of the  
15 genes (Yang, J. et al. (1995) J. Biol. Chem. 270:12152-12161). Regulation of cholesterol uptake and biosynthesis also occurs via the oxysterol-binding protein (OSBP). Oxysterols are oxidation products formed during the catabolism of cholesterol, and are involved in regulation of steroid biosynthesis. OSBP is a high-affinity intracellular receptor for a variety of oxysterols that down-regulate cholesterol synthesis and stimulate cholesterol esterification (Lagace, T.A. et al. (1997) Biochem. J. 326:205-  
20 213).

Supernatant protein factor (SPF), which stimulates squalene epoxidation and conversion of squalene to lanosterol, is a cytosolic squalene transfer protein that enhances cholesterol biosynthesis. Squalene epoxidase, a membrane-associated enzyme that converts squalene to squalene 2,3-oxide, plays an important role in the maintenance of cholesterol homeostasis. SPF belongs to a family of  
25 cytosolic lipid-binding/transfer proteins such as alpha-tocopherol transfer protein, cellular retinal binding protein, yeast phosphatidylinositol transfer protein (Sec14p), and squid retinal binding protein (Shibata, N. et al. (2001) Proc. Natl. Acad. Sci. USA 98:2244-2249).

#### Lipid Metabolism Enzymes

Long-chain fatty acids are also substrates for eicosanoid production, and are important in the  
30 functional modification of certain complex carbohydrates and proteins. 16-carbon and 18-carbon fatty acids are the most common. Fatty acid synthesis occurs in the cytoplasm. In the first step, acetyl-Coenzyme A (CoA) carboxylase (ACC) synthesizes malonyl-CoA from acetyl-CoA and bicarbonate. The enzymes which catalyze the remaining reactions are covalently linked into a single polypeptide

chain, referred to as the multifunctional enzyme fatty acid synthase (FAS). FAS catalyzes the synthesis of palmitate from acetyl-CoA and malonyl-CoA. FAS contains acetyl transferase, malonyl transferase,  $\beta$ -ketoacetyl synthase, acyl carrier protein,  $\beta$ -ketoacyl reductase, dehydratase, enoyl reductase, and thioesterase activities. The final product of the FAS reaction is the 16-carbon fatty acid palmitate. Further elongation, as well as unsaturation, of palmitate by accessory enzymes of the ER produces the variety of long chain fatty acids required by the individual cell. These enzymes include a NADH-cytochrome  $b_5$  reductase, cytochrome  $b_5$ , and a desaturase.

Within cells, fatty acids are transported by cytoplasmic fatty acid binding proteins (Online Mendelian Inheritance in Man (OMIM) \*134650 Fatty Acid-Binding Protein 1, Liver; FABP1). Diazepam binding inhibitor (DBI), also known as endozepine and acyl CoA-binding protein, is an endogenous  $\gamma$ -aminobutyric acid (GABA) receptor ligand which is thought to down-regulate the effects of GABA. DBI binds medium- and long-chain acyl-CoA esters with very high affinity and may function as an intracellular carrier of acyl-CoA esters (OMIM \*125950 Diazepam Binding Inhibitor; DBI; PROSITE PDOC00686 Acyl-CoA-binding protein signature).

Fat stored in liver and adipose triglycerides may be released by hydrolysis and transported in the blood. Free fatty acids are transported in the blood by albumin. Triacylglycerols, also known as triglycerides and neutral fats, are major energy stores in animals. Triacylglycerols are esters of glycerol with three fatty acid chains. Glycerol-3-phosphate is produced from dihydroxyacetone phosphate by the enzyme glycerol phosphate dehydrogenase or from glycerol by glycerol kinase. Fatty acid-CoAs are produced from fatty acids by fatty acyl-CoA synthetases. Glycerol-3-phosphate is acylated with two fatty acyl-CoAs by the enzyme glycerol phosphate acyltransferase to give phosphatidate. Phosphatidate phosphatase converts phosphatidate to diacylglycerol, which is subsequently acylated to a triacylglycerol by the enzyme diglyceride acyltransferase. Phosphatidate phosphatase and diglyceride acyltransferase form a triacylglycerol synthetase complex bound to the ER membrane.

Mitochondrial and peroxisomal beta-oxidation enzymes degrade saturated and unsaturated fatty acids by sequential removal of two-carbon units from CoA-activated fatty acids. The main beta-oxidation pathway degrades both saturated and unsaturated fatty acids while the auxiliary pathway performs additional steps required for the degradation of unsaturated fatty acids. The pathways of mitochondrial and peroxisomal beta-oxidation use similar enzymes, but have different substrate specificities and functions. Mitochondria oxidize short-, medium-, and long-chain fatty acids to produce energy for cells. Mitochondrial beta-oxidation is a major energy source for cardiac and skeletal muscle. In liver, it provides ketone bodies to the peripheral circulation when glucose levels are

low as in starvation, endurance exercise, and diabetes (Eaton, S. et al. (1996) *Biochem. J.* 320:345-357). Peroxisomes oxidize medium-, long-, and very-long-chain fatty acids, dicarboxylic fatty acids, branched fatty acids, prostaglandins, xenobiotics, and bile acid intermediates. The chief roles of peroxisomal beta-oxidation are to shorten toxic lipophilic carboxylic acids to facilitate their excretion and to shorten very-long-chain fatty acids prior to mitochondrial beta-oxidation (Mannaerts, G.P. and P.P. Van Veldhoven (1993) *Biochimie* 75:147-158). Enzymes involved in beta-oxidation include acyl CoA synthetase, carnitine acyltransferase, acyl CoA dehydrogenases, enoyl CoA hydratases, L-3-hydroxyacyl CoA dehydrogenase,  $\beta$ -ketothiolase, 2,4-dienoyl CoA reductase, and isomerase.

Three classes of lipid metabolism enzymes are discussed in further detail. The three classes are lipases, phospholipases and lipoxigenases.

### Lipases

Triglycerides are hydrolyzed to fatty acids and glycerol by lipases. Adipocytes contain lipases that break down stored triacylglycerols, releasing fatty acids for export to other tissues where they are required as fuel. Lipases are widely distributed in animals, plants, and prokaryotes. Triglyceride lipases (ExPASy ENZYME EC 3.1.1.3), also known as triacylglycerol lipases and tributyrases, hydrolyze the ester bond of triglycerides. In higher vertebrates there are at least three tissue-specific isozymes including gastric, hepatic, and pancreatic lipases. These three types of lipases are structurally closely related to each other as well as to lipoprotein lipase. The most conserved region in gastric, hepatic, and pancreatic lipases is centered around a serine residue which is also present in lipases of prokaryotic origin. Mutation in the serine residue renders the enzymes inactive. Gastric, hepatic, and pancreatic lipases hydrolyze lipoprotein triglycerides and phospholipids. Gastric lipases in the intestine aid in the digestion and absorption of dietary fats. Hepatic lipases are bound to and act at the endothelial surfaces of hepatic tissues. Hepatic lipases also play a major role in the regulation of plasma lipids. Pancreatic lipase requires a small protein cofactor, colipase, for efficient dietary lipid hydrolysis. Colipase binds to the C-terminal, non-catalytic domain of lipase, thereby stabilizing an active conformation and considerably increasing the overall hydrophobic binding site. Deficiencies of these enzymes have been identified in man, and all are associated with pathologic levels of circulating lipoprotein particles (Gargouri, Y. et al. (1989) *Biochim. Biophys. Acta* 1006:255-271; Connelly, P.W. (1999) *Clin. Chim. Acta* 286:243-255; van Tilbeurgh, H. et al. (1999) *Biochim. Biophys. Acta* 1441:173-184).

Lipoprotein lipases (ExPASy ENZYME EC 3.1.1.34), also known as clearing factor lipases, diglyceride lipases, or diacylglycerol lipases, hydrolyze triglycerides and phospholipids present in circulating plasma lipoproteins, including chylomicrons, very low and intermediate density lipoproteins



and high-density lipoproteins (HDL). Together with pancreatic and hepatic lipases, lipoprotein lipases (LPL) share a high degree of primary sequence homology. Both lipoprotein lipases and hepatic lipases are anchored to the capillary endothelium via glycosaminoglycans and can be released by intravenous administration of heparin. LPLs are primarily synthesized by adipocytes, muscle cells, and macrophages. Catalytic activities of LPLs are activated by apolipoprotein C-II and are inhibited by high ionic strength conditions such as 1 M NaCl. LPL deficiencies in humans contribute to metabolic diseases such as hypertriglyceridemia, HDL2 deficiency, and obesity (Jackson, R.L. (1983) in The Enzymes (Boyer, P.D., ed.) Vol. XVI, pp. 141-186, Academic Press, New York NY; Eckel, R.H. (1989) *New Engl. J. Med.* 320:1060-1068).

10 Phospholipases

Phospholipases, a group of enzymes that catalyze the hydrolysis of membrane phospholipids, are classified according to the bond cleaved in a phospholipid. They are classified into PLA1, PLA2, PLB, PLC, and PLD families. Phospholipases are involved in many inflammatory reactions by making arachidonate available for eicosanoid biosynthesis. More specifically, arachidonic acid is processed into bioactive lipid mediators of inflammation such as lyso-platelet-activating factor and eicosanoids. The synthesis of arachidonic acid from membrane phospholipids is the rate-limiting step in the biosynthesis of the four major classes of eicosanoids (prostaglandins, prostacyclins, thromboxanes and leukotrienes), which are 20-carbon molecules derived from fatty acids. Eicosanoids are signaling molecules which have roles in pain, fever, and inflammation. The precursor of all eicosanoids is arachidonate, which is generated from phospholipids by phospholipase A<sub>2</sub> and from diacylglycerols by diacylglycerol lipase. Leukotrienes are produced from arachidonate by the action of lipoxygenases (Kaiser, E. et al. (1990) *Clin. Biochem.* 23:349-370). Furthermore, leukotriene-B<sub>4</sub> is known to function in a feedback loop which further increases PLA2 activity (Wijkander, J. et al. (1995) *J. Biol. Chem.* 270:26543-26549).

25 The secretory phospholipase A<sub>2</sub> (PLA2) superfamily comprises a number of heterogeneous enzymes whose common feature is to hydrolyze the sn-2 fatty acid acyl ester bond of phosphoglycerides. Hydrolysis of the glycerophospholipids releases free fatty acids and lysophospholipids. PLA2 activity generates precursors for the biosynthesis of biologically active lipids, hydroxy fatty acids, and platelet-activating factor. PLA2s were first described as components of snake venoms, and were later characterized in numerous species. PLA2s have traditionally been classified into several major groups and subgroups based on their amino acid sequences, divalent cation requirements, and location of disulfide bonds. The PLA2s of Groups I, II, and III consist of low molecular weight, secreted, Ca<sup>2+</sup>-dependent proteins. Group IV PLA2s are primarily 85-kDa,

Ca<sup>2+</sup>-dependent cytosolic phospholipases. Finally, a number of Ca<sup>2+</sup>-independent PLA2s have been described, which comprise Group V (Davidson, F.F. and E.A. Dennis (1990) *J. Mol. Evol.* 31:228-238; and Dennis, E.F. (1994) *J. Biol. Chem.* 269:13057-13060).

The first PLA2s to be extensively characterized were the Group I, II, and III PLA2s found in snake and bee venoms. These venom PLA2s share many features with mammalian PLA2s including a common catalytic mechanism, the same Ca<sup>2+</sup> requirement, and conserved primary and tertiary structures. In addition to their role in the digestion of prey, the venom PLA2s display neurotoxic, myotoxic, anticoagulant, and proinflammatory effects in mammalian tissues. This diversity of pathophysiological effects is due to the presence of specific, high affinity receptors for these enzymes on various cells and tissues (Lambeau, G. et al. (1995) *J. Biol. Chem.* 270:5534-5540).

PLA2s from Groups I, IIA, IIC, and V have been described in mammalian and avian cells, and were originally characterized by tissue distribution, although the distinction is no longer absolute. Thus, Group I PLA2s were found in the pancreas, Group IIA and IIC were derived from inflammation-associated tissues (e.g., the synovium), and Group V were from cardiac tissue. The pancreatic PLA2s function in the digestion of dietary lipids and have been proposed to play a role in cell proliferation, smooth muscle contraction, and acute lung injury. The Group II inflammatory PLA2s are potent mediators of inflammatory processes and are highly expressed in serum and synovial fluids of patients with inflammatory disorders. These Group II PLA2s are found in most human cell types assayed and are expressed in diverse pathological processes such as septic shock, intestinal cancers, rheumatoid arthritis, and epidermal hyperplasia. A Group V PLA2 has been cloned from brain tissue and is strongly expressed in heart tissue. A human PLA2 was recently cloned from fetal lung, and based on its structural properties, appears to be the first member of a new group of mammalian PLA2s, referred to as Group X. Other PLA2s have been cloned from various human tissues and cell lines, suggesting a large diversity of PLA2s (Chen, J. et al. (1994) *J. Biol. Chem.* 269:2365-2368; Kennedy, B.P. et al. (1995) *J. Biol. Chem.* 270: 22378-22385; Komada, M. et al. (1990) *Biochem. Biophys. Res. Commun.* 168:1059-1065; Cupillard, L. et al. (1997) *J. Biol. Chem.* 272:15745-15752; and Nalefski, E.A. et al. (1994) *J. Biol. Chem.* 269:18239-18249).

Phospholipases B (PLB) (ExPASy ENZYME EC 3.1.1.5), also known as lysophospholipase, lecithinase B, or lysolecithinase are widely distributed enzymes that metabolize intracellular lipids, and occur in numerous isoforms. Small isoforms, approximately 15-30 kD, function as hydrolases; large isoforms, those exceeding 60 kD, function both as hydrolases and transacylases. A particular substrate for PLBs, lysophosphatidylcholine, causes lysis of cell membranes when it is formed or imported into a cell. PLBs are regulated by lipid factors including acylcarnitine, arachidonic acid, and

phosphatidic acid. These lipid factors are signaling molecules important in numerous pathways, including the inflammatory response (Anderson, R. et al. (1994) *Toxicol. Appl. Pharmacol.* 125:176-183; Selle, H. et al. (1993); *Eur. J. Biochem.* 212:411-416).

Phospholipase C (PLC) (ExPASy ENZYME EC 3.1.4.10) plays an important role in  
5 transmembrane signal transduction. Many extracellular signaling molecules including hormones, growth factors, neurotransmitters, and immunoglobulins bind to their respective cell surface receptors and activate PLCs. The role of an activated PLC is to catalyze the hydrolysis of phosphatidyl-inositol-4,5-bisphosphate (PIP<sub>2</sub>), a minor component of the plasma membrane, to produce diacylglycerol and inositol 1,4,5-trisphosphate (IP<sub>3</sub>). In their respective biochemical pathways, IP<sub>3</sub>  
10 and diacylglycerol serve as second messengers and trigger a series of intracellular responses. IP<sub>3</sub> induces the release of Ca<sup>2+</sup> from internal cellular storage, and diacylglycerol activates protein kinase C (PKC). Both pathways are part of transmembrane signal transduction mechanisms which regulate cellular processes which include secretion, neural activity, metabolism, and proliferation.

Several distinct isoforms of PLC have been identified and are categorized as PLC-beta,  
15 PLC-gamma, and PLC-delta. Subtypes are designated by adding Arabic numbers after the Greek letters, eg. PLC-B-1. PLCs have a molecular mass of 62-68 kDa, and their amino acid sequences show two regions of significant similarity. The first region, designated X, has about 170 amino acids, and the second, or Y region, contains about 260 amino acids.

The catalytic activities of the three isoforms of PLC are dependent upon Ca<sup>2+</sup>. It has been  
20 suggested that the binding sites for Ca<sup>2+</sup> in the PLCs are located in the Y-region, one of two conserved regions. The hydrolysis of common inositol-containing phospholipids, such as phosphatidylinositol (PI), phosphatidylinositol 4-monophosphate (PIP), and phosphatidylinositol 4,5-bisphosphate (PIP<sub>2</sub>), by any of the isoforms yields cyclic and noncyclic inositol phosphates (Rhee, S.G. and Y.S. Bae (1997) *J. Biol. Chem.* 272:15045-15048).

25 All mammalian PLCs contain a pleckstrin homology (PH) domain which is about 100 amino acids in length and is composed of two antiparallel beta sheets flanked by an amphipathic alpha helix. PH domains target PLCs to the membrane surface by interacting with either the beta/gamma subunits of G proteins or PIP<sub>2</sub> (PROSITE PDOC50003).

Phospholipase D (PLD) (ExPASy ENZYME EC 3.1.4.4), also known as lecithinase D,  
30 lipophosphodiesterase II, and choline phosphatase catalyzes the hydrolysis of phosphatidylcholine and other phospholipids to generate phosphatidic acid. PLD plays an important role in membrane vesicle trafficking, cytoskeletal dynamics, and transmembrane signal transduction. In addition, the activation of PLD is involved in cell differentiation and growth (reviewed in Liscovitch, M. (2000) *Biochem. J.*

345:401-415).

PLD is activated in mammalian cells in response to diverse stimuli that include hormones, neurotransmitters, growth factors, cytokines, activators of protein kinase C, and agonist binding to G-protein-coupled receptors. At least two forms of mammalian PLD, PLD1 and PLD2, have been  
5 identified. PLD1 is activated by protein kinase C alpha and by the small GTPases ARF and RhoA. (Houle, M.G. and S. Bourgoin (1999) *Biochim. Biophys. Acta* 1439:135-149). PLD2 can be selectively activated by unsaturated fatty acids such as oleate (Kim, J.H. (1999) *FEBS Lett.* 454:42-46).

#### Lipoxygenases

10 Lipoxygenases (ExPASy ENZYME EC 1.13.11.12) are non-heme iron-containing enzymes that catalyze the dioxygenation of certain polyunsaturated fatty acids such as lipoproteins. Lipoxygenases are found widely in plants, fungi, and animals. Several different lipoxygenase enzymes are known, each having a characteristic oxidation action. In animals, there are specific lipoxygenases that catalyze the dioxygenation of arachidonic acid at the carbon-3, 5, 8, 11, 12, and 15 positions.  
15 These enzymes are named after the position of arachidonic acid that they dioxygenate. Lipoxygenases have a single polypeptide chain with a molecular mass of ~75-80 kDa in animals. The proteins have an N-terminal-barrel domain and a larger catalytic domain containing a single atom of non-heme iron. Oxidation of the ferric enzyme to an active form is required for catalysis (Yamamoto, S. (1992) *Biochim. Biophys. Acta* 1128:117-131; Brash, A.R. (1999) *J. Biol. Chem.* 274:23679-23682). A variety of lipoxygenase inhibitors exist and are classified into five major  
20 categories according to their mechanism of inhibition. These include antioxidants, iron chelators, substrate analogues, lipoxygenase-activating protein inhibitors, and, finally, epidermal growth factor-receptor inhibitors.

3-Lipoxygenase, also known as e-LOX-3 or Aloxe3 has recently been cloned from murine  
25 epidermis. Aloxe3 resides on mouse chromosome 11, and the deduced amino acid sequence for Aloxe3 is 54% identical to the 12-lipoxygenase sequences (Kinzig, A. (1999) *Genomics* 58:158-164).

5-Lipoxygenase (5-LOX, ExPASy ENZYME EC 1.13.11.34), also known as arachidonate:oxygen 5-oxidoreductase, is found primarily in white blood cells, macrophages, and mast cells. 5-LOX converts arachidonic acid first to 5-hydroperoxyeicosatetraenoic acid (5-HPETE) and  
30 then to leukotriene (LTA4 (5,6-oxido-7,9,11,14-eicosatetraenoic acid)). Subsequent conversion of leukotriene A4 by leukotriene A4 hydrolase yields the potent neutrophil chemoattractant leukotriene B4. Alternatively, conjugation of LTA4 with glutathione by leukotriene C4 synthase plus downstream metabolism leads to the cysteinyl leukotrienes that influence airway reactivity and mucus secretion,

especially in asthmatics. Most lipoxygenases require no other cofactors or proteins for activity. In contrast, the mammalian 5-LOX requires calcium and ATP, and is activated in the presence of a 5-LOX activating protein (FLAP). FLAP itself binds to arachidonic acid and supplies 5-LOX with substrate (Lewis, R.A. et al. (1990) *New Engl. J. Med.* 323:645-655). The expression levels of 5-LOX and FLAP are found to be increased in the lungs of patients with plexogenic (primary) pulmonary hypertension (Wright, L. et al. (1998) *Am. J. Respir. Crit. Care Med.* 157:219-229).

12-Lipoxygenase (12-LOX, ExPASy ENZYME: EC 1.13.11.31) oxygenates arachidonic acid to form 12-hydroperoxyeicosatetraenoic acid (12-HPETE). Mammalian 12-lipoxygenases are named after the prototypical tissues of their occurrence (hence, the leukocyte, platelet, or epidermal types).

10 Platelet-type 12-LOX has been found to be the predominant isoform in epidermal skin specimens and epidermoid cells. Leukocyte 12-LOX was first characterized extensively from porcine leukocytes and was found to have a rather broad distribution in mammalian tissues by immunochemical assays. Besides tissue distribution, the leukocyte 12-LOX is distinguished from the platelet-type enzyme by its ability to form 15-HPETE, in addition to 12-HPETE, from arachidonic acid substrate. Leukocyte 12-LOX is highly related to 15-lipoxygenase (15-LOX) in that both are dual specificity lipoxygenases, and they are about 85% identical in primary structure in higher mammals. Leukocyte 12-LOX is found in tracheal epithelium, leukocytes, and macrophages (Conrad, D.J. (1999) *Clin. Rev. Allergy Immunol.* 17:71-89).

15-Lipoxygenase (15-LOX; ExPASy ENZYME: EC 1.13.11.33) is found in human reticulocytes, airway epithelium, and eosinophils. 15-LOX has been detected in atherosclerotic lesions in mammals, specifically rabbit and man. The enzyme, in addition to its role in oxidative modification of lipoproteins, is important in the inflammatory reaction in atherosclerotic lesions. 15-LOX has been shown to be induced in human monocytes by the cytokine IL-4, which is known to be implicated in the inflammatory process (Kuhn, H. and S. Borngraber (1999) *Adv. Exp. Med. Biol.* 447:5-28).

25 A variety of lipolytic enzymes with a GDSL-like motif as part of the active site have been identified. Members of this family include a lipase/acylhydrolase, thermolabile hemolysin and rabbit phospholipase (AdRab-B)(Interpro entry IPR001087, <http://www.sanger.ac.uk>). A homolog of AdRab-B is guinea pig intestinal phospholipase B, a calcium-independent phospholipase that contributes to lipid digestion as an ectoenzyme by sequentially hydrolyzing the acyl ester bonds of glycerophospholipids. Phospholipase B also has a role in male reproduction (Delagebeaudeuf, C. et al. (1998) *J. Biol. Chem.* 273:13407-13414).

#### Lipid-Associated Molecules and Disease

Lipids and their associated proteins have roles in human diseases and disorders. Increased

synthesis of long-chain fatty acids occurs in neoplasms including those of the breast, prostate, ovary, colon and endometrium.

In the arterial disease atherosclerosis, fatty lesions form on the inside of the arterial wall. These lesions promote the loss of arterial flexibility and the formation of blood clots (Guyton, supra).

- 5 There is a strong inverse correlation between the levels of plasma HDL and risk of premature coronary heart disease. Absence of the LDL receptor, the cause of familial hypercholesterolemia, leads to increased plasma cholesterol levels and ultimately to atherosclerosis (Stryer, supra, pp. 691-702). Oxysterols are present in human atherosclerotic plaques and are believed to play an active role in plaque development (Brown, A.J. (1999) *Atherosclerosis* 142:1-28). Lipases, phospholipases, and
- 10 lipoxygenases are thought to contribute to complex diseases, such as atherosclerosis, obesity, arthritis, asthma, and cancer, as well as to single gene defects, such as Wolman's disease and Type I hyperlipoproteinemia.

- Steatosis, or fatty liver, is characterized by the accumulation of triglycerides in the liver and may occur in association with a variety of conditions including alcoholism, diabetes, obesity, and
- 15 prolonged parenteral nutrition. Steatosis may lead to fibrosis and cirrhosis of the liver.

- Niemann-Pick diseases types A and B are caused by accumulation of sphingomyelin (a sphingolipid) and other lipids in the central nervous system due to a defect in the enzyme sphingomyelinase, leading to neurodegeneration and lung disease. Niemann-Pick disease type C results from a defect in cholesterol transport, leading to the accumulation of sphingomyelin and
- 20 cholesterol in lysosomes and a secondary reduction in sphingomyelinase activity. Neurological symptoms such as grand mal seizures, ataxia, and loss of previously learned speech, manifest 1-2 years after birth. A mutation in the NPC protein, which contains a putative cholesterol-sensing domain, was found in a mouse model of Niemann-Pick disease type C (Fauci, supra, p. 2175; Loftus, S.K. et al. (1997) *Science* 277:232-235).

- 25 Tay-Sachs disease is an autosomal recessive, progressive neurodegenerative disorder caused by the accumulation of the G<sub>M2</sub> ganglioside in the brain (Igdoura, S.A. et al. (1999) *Hum. Mol. Genet.* 8:1111-6) due to a deficiency of the enzyme hexosaminidase A. The disease is characterized by the onset of developmental retardation, followed by paralysis, dementia, blindness, and usually death within the second or third year of life. Confirmatory evidence of Tay-Sachs disease is obtained at autopsy
- 30 upon the identification of ballooned neurons in the central nervous system (Online Mendelian Inheritance in Man (OMIM). Johns Hopkins University, Baltimore, MD. MIM Number: 272800, 8/4/2000, WWW URL: <http://www.ncbi.nlm.nih.gov/omim/>). In the case of Tay-Sachs disease, cortical pyramidal neurons undergo a second round of dendritogenesis (Walkley, S.U. et al. (1998)

Ann. N.Y. Acad. Sci. 845:188-99).

Other diseases are also associated with defects in sialidase activity.  $G_{M1}$  gangliosidosis and Morquio B disease both arise from beta-galactosidase deficiency, although the diseases present with distinct phenotypes. Sialidosis arises from a neuraminidase deficiency but presents with symptoms  
5 similar to gangliosidosis. A likely reason for the overlapping phenotypes of sialidase deficiencies is the presence of these enzymes in a complex in lysosomes (Callahan, J.W. (1999) Biochim. Biophys. Acta 1455:85-103).

PLAs are implicated in a variety of disease processes. For example, PLAs are found in the pancreas, in cardiac tissue, and in inflammation-associated tissues. Pancreatic PLAs function in the  
10 digestion of dietary lipids and have been proposed to play a role in cell proliferation, smooth muscle contraction, and acute lung injury. Inflammatory PLAs are potent mediators of inflammatory processes and are highly expressed in serum and synovial fluids of patients with inflammatory disorders. Additionally, inflammatory PLAs are found in most human cell types and are expressed in diverse pathological processes such as septic shock, intestinal cancers, rheumatoid arthritis, and  
15 epidermal hyperplasia.

The role of PLBs in human tissues has been investigated in various research studies. Hydrolysis of lysophosphatidylcholine by PLBs causes lysis in erythrocyte membranes (Selle, supra). Similarly, Endresen, M.J. et al. (1993; Scand. J. Clin. Invest. 53:733-739) reported that the increased  
20 hydrolysis of lysophosphatidylcholine by PLB in pre-eclamptic women causes release of free fatty acids into the sera. In renal studies, PLB was shown to protect  $Na^+, K^+$ -ATPase from the cytotoxic and cytolytic effects of cyclosporin A (Anderson, supra).

Lipases, phospholipases, and lipoxygenases are thought to contribute to complex diseases, such as atherosclerosis, obesity, arthritis, asthma, and cancer, as well as to single gene defects, such as Wolman's disease and Type I hyperlipoproteinemia.

25 The discovery of new lipid-associated molecules, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cancers, neurological, autoimmune/inflammatory, gastrointestinal, and cardiovascular disorders, and disorders of lipid metabolism, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of lipid-associated molecules.

30

### SUMMARY OF THE INVENTION

The invention features purified polypeptides, lipid-associated molecules, referred to collectively as "LIPAM" and individually as "LIPAM-1," "LIPAM-2," "LIPAM-3," "LIPAM-4," "LIPAM-5,"

“LIPAM-6,” “LIPAM-7,” “LIPAM-8,” and “LIPAM-9.” In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-9.

10 The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-9. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:10-18.

20 Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

25 The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group



consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide  
5 encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the  
10 group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

The invention further provides an isolated polynucleotide selected from the group consisting of  
15 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide  
20 comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least  
25 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to  
30 said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional LIPAM, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional LIPAM, comprising

administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a polypeptide  
5 comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) exposing a sample comprising the  
10 polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional LIPAM, comprising administering to a patient in need of such treatment the composition.

15 The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having  
20 an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

25 The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, c) a biologically active fragment of a polypeptide having  
30 an amino acid sequence selected from the group consisting of SEQ ID NO:1-9, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-9. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity

of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

5           The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target  
10 polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

          The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20  
15 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of  
20 ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a  
25 polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated  
30 biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now

described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## 5 DEFINITIONS

"LIPAM" refers to the amino acid sequences of substantially purified LIPAM obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of  
10 LIPAM. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of LIPAM either by directly interacting with LIPAM or by acting on components of the biological pathway in which LIPAM participates.

An "allelic variant" is an alternative form of the gene encoding LIPAM. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in  
15 polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

20 "Altered" nucleic acid sequences encoding LIPAM include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as LIPAM or a polypeptide with at least one functional characteristic of LIPAM. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding LIPAM, and improper or unexpected hybridization to allelic variants, with  
25 a locus other than the normal chromosomal locus for the polynucleotide sequence encoding LIPAM. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent LIPAM. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological  
30 or immunological activity of LIPAM is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains

having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of LIPAM. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of LIPAM either by directly interacting with LIPAM or by acting on components of the biological pathway in which LIPAM participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind LIPAM polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No.

5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH<sub>2</sub>), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

10       The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

      The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed  
15       nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

      The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as  
20       phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring  
25       nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

      The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic"  
30       refers to the capability of the natural, recombinant, or synthetic LIPAM, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

      "Complementary" describes the relationship between two single-stranded nucleic acid



sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding LIPAM or fragments of LIPAM may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
25	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
30	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
35	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr

Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

---

5

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

10 A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains  
 15 at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

20 "Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an  
 25 exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of LIPAM or the polynucleotide encoding LIPAM which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up  
 30 to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For  
 35 example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected

from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

5 A fragment of SEQ ID NO:10-18 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:10-18, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:10-18 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:10-18 from related polynucleotide sequences. The precise length of a fragment of SEQ ID  
10 NO:10-18 and the region of SEQ ID NO:10-18 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-9 is encoded by a fragment of SEQ ID NO:10-18. A fragment of SEQ ID NO:1-9 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-9. For example, a fragment of SEQ ID NO:1-9 is useful as an immunogenic peptide for the  
15 development of antibodies that specifically recognize SEQ ID NO:1-9. The precise length of a fragment of SEQ ID NO:1-9 and the region of SEQ ID NO:1-9 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon  
20 (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to  
25 the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default  
30 parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS

8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

- 5 Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis
- 10 programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST
- 15 programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

20 *Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

25 *Filter: on*

- Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous
- 30 nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode

similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances,

such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

5       The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate  
10 to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression  
15 of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of LIPAM which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment  
20 of LIPAM which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other  
25 chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of LIPAM. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of LIPAM.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide,  
30 polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a

functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

5           “Peptide nucleic acid” (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

10           “Post-translational modification” of an LIPAM may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of LIPAM.

          “Probe” refers to nucleic acid sequences encoding LIPAM, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. “Primers” are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

          Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

          Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that



purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to  
5 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase  
10 sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may  
15 also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved  
20 oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

25 A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have  
30 been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing LIPAM, nucleic acids encoding LIPAM, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters,

chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression  
5 by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based  
10 on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals  
15 and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a  
20 recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation.  
25 Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-  
30 1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an

“allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass “single nucleotide polymorphisms” (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A “variant” of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## 20 THE INVENTION

The invention is based on the discovery of new human lipid-associated molecules (LIPAM), the polynucleotides encoding LIPAM, and the use of these compositions for the diagnosis, treatment, or prevention of cancers, neurological, autoimmune/inflammatory, gastrointestinal, and cardiovascular disorders, and disorders of lipid metabolism.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the

polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s).  
5 Column 5 shows the annotation of the GenBank homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column  
10 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable  
15 databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are lipid-associated molecules. For example, SEQ ID NO:1 is 43% identical, from residue M336 to residue R989, to human cytosolic phospholipase A2 beta (cPLA2beta) (GenBank ID g4886978) as determined by the Basic Local Alignment Search Tool  
20 (BLAST). (See Table 2.) The BLAST probability score is  $1.6e-161$ , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. In an alternative example, SEQ ID NO:3 is 41% identical, from residue M1 to residue Y644, to rat phospholipase C delta-4 (GenBank ID g4894788) as determined by BLAST, with a probability score of  $2.5e-126$ . (See Table 2.) SEQ ID NO:3 also contains phosphatidylinositol-specific phospholipase C, X and Y domains and a C2  
25 domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analyses provide further corroborative evidence that SEQ ID NO:3 is a phospholipase C. In an alternative example, SEQ ID NO:5 is 40% identical, from residue M1 to residue N316, to human phosphatidylserine-specific phospholipase A1 (GenBank ID g4090960) as determined by BLAST, with  
30 a probability score of  $3.1e-63$ . (See Table 2.) SEQ ID NO:5 also contains a lipase domain as determined by searching for statistically significant matches in the HMM-based PFAM database. (See Table 3.) Data from BLIMPS analyses provide further corroborative evidence that SEQ ID NO:5 is a lipase. In an alternative example, SEQ ID NO:6 is 45% identical from residue C41 to

residue I491, 39% identical from residue K582 to residue K899, 33% identical from residue S541 to residue G603, and 22% identical from residue S519 to residue L571, to Mus musculus phospholipase C-L2 (GenBank ID g6705987) as determined by BLAST, with probability scores of  $1.1\text{e-}164$  from residue C41 to residue I491,  $1.1\text{e-}164$  from residue K582 to residue K899,  $1.0\text{e-}56$  from residue S541 to residue G603, and  $1.1\text{e-}55$  from residue S519 to residue L571. (See Table 2.) SEQ ID NO:6 also contains phosphatidylinositol-specific phospholipase active site domains as determined by searching for statistically significant matches in the HMM-based PFAM database. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:6 is a phospholipase. In an alternative example, SEQ ID NO:7 is 90% identical, from residue M1 to residue K1294, to Mus musculus M-RdgB2 retinal degeneration protein B subtype 2 (GenBank ID g5771350) as determined by BLAST, with a probability score of 0.0. (See Table 2.) SEQ ID NO:7 also contains a phosphatidylinositol transfer protein domain as determined by searching for statistically significant matches in the HMM-based PFAM database. (See Table 3.) Data from BLIMPS, MOTIFS, and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:7 is a phosphatidylinositol transfer protein. In an alternative example, SEQ ID NO:9 is 81% identical from residue R387 to residue T546, 69% identical from residue T181 to residue Q358, and 61% identical from residue A2 to residue D191, to Mus musculus TAGL-beta (GenBank ID g6651241) as determined by BLAST, with a probability score of  $3.8\text{e-}188$ . (See Table 2.) Data from additional BLAST analyses provide further corroborative evidence that SEQ ID NO:9 is a protein peptidoglycan recognition precursor. SEQ ID NO:2, SEQ ID NO:4, and SEQ ID NO:8 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-9 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide sequences of the invention, and of fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:10-18 or that distinguish between SEQ ID NO:10-18 and related polynucleotide sequences.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for

example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the

5 ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in

10 column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3...</sub>, if present, represent specific exons

15 that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human

20 genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

25 Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.

FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in  
 5 Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences  
 10 which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses LIPAM variants. A preferred LIPAM variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the LIPAM amino acid sequence, and which contains at least one functional or  
 15 structural characteristic of LIPAM.

The invention also encompasses polynucleotides which encode LIPAM. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:10-18, which encodes LIPAM. The polynucleotide sequences of SEQ ID NO:10-18, as presented in the Sequence Listing, embrace the equivalent RNA  
 20 sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding LIPAM. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide  
 25 sequence encoding LIPAM. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:10-18 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:10-18. Any one of the polynucleotide variants described above can encode an amino acid  
 30 sequence which contains at least one functional or structural characteristic of LIPAM.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a



polynucleotide sequence encoding LIPAM. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding LIPAM, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%,  
5 or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to the polynucleotide sequence encoding LIPAM over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide sequence encoding LIPAM. Any one of the splice variants described above can  
10 encode an amino acid sequence which contains at least one functional or structural characteristic of LIPAM.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding LIPAM, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be  
15 produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring LIPAM, and all such variations are to be considered as being specifically disclosed.

20 Although nucleotide sequences which encode LIPAM and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring LIPAM under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding LIPAM or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the  
25 peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding LIPAM and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

30 The invention also encompasses production of DNA sequences which encode LIPAM and LIPAM derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce

mutations into a sequence encoding LIPAM or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:10-18 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and  
5 S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment  
10 of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV),  
15 PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M.  
20 (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding LIPAM may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed,  
25 restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids  
30 Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown

sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060).

Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in  
5 finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

10 When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

15 Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate  
20 software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which  
25 encode LIPAM may be cloned in recombinant DNA molecules that direct expression of LIPAM, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express LIPAM.

The nucleotide sequences of the present invention can be engineered using methods generally  
30 known in the art in order to alter LIPAM-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-

mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No.

5 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of LIPAM, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to  
10 selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively,  
15 fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding LIPAM may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids  
20 Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, LIPAM itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved  
25 using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of LIPAM, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid  
30 chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active LIPAM, the nucleotide sequences encoding LIPAM

or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in  
5 polynucleotide sequences encoding LIPAM. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding LIPAM. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding LIPAM and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional  
10 or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See,  
15 e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding LIPAM and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A  
20 Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding LIPAM. These include, but are not limited to, microorganisms such as bacteria transformed  
25 with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster  
30 (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and

Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding LIPAM. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding LIPAM can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding LIPAM into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of LIPAM are needed, e.g. for the production of antibodies, vectors which direct high level expression of LIPAM may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of LIPAM. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of LIPAM. Transcription of sequences encoding LIPAM may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated

transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding LIPAM may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses LIPAM in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of LIPAM in cell lines is preferred. For example, sequences encoding LIPAM can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk<sup>-</sup>* and *apr<sup>-</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which

alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to  
5 quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding LIPAM is inserted within a marker gene sequence, transformed cells  
10 containing sequences encoding LIPAM can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding LIPAM under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding LIPAM and that  
15 express LIPAM may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of LIPAM using either  
20 specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on LIPAM is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See,  
25 e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and  
30 may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding LIPAM include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding LIPAM, or any fragments thereof, may be cloned into a vector



for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega  
5 (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding LIPAM may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein  
10 produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode LIPAM may be designed to contain signal sequences which direct secretion of LIPAM through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the  
15 inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities  
20 (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding LIPAM may be ligated to a heterologous sequence resulting in translation of a  
25 fusion protein in any of the aforementioned host systems. For example, a chimeric LIPAM protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of LIPAM activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose  
30 binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion

proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the LIPAM encoding sequence and the heterologous protein sequence, so that LIPAM may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled LIPAM may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

LIPAM of the present invention or fragments thereof may be used to screen for compounds that specifically bind to LIPAM. At least one and up to a plurality of test compounds may be screened for specific binding to LIPAM. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of LIPAM, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which LIPAM binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express LIPAM, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing LIPAM or cell membrane fractions which contain LIPAM are then contacted with a test compound and binding, stimulation, or inhibition of activity of either LIPAM or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with LIPAM, either in solution or affixed to a solid support, and detecting the binding of LIPAM to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

LIPAM of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of LIPAM. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for LIPAM activity, wherein LIPAM is combined with at least one test compound, and the activity of LIPAM in the presence of a test compound is compared with the activity of LIPAM in the absence of the test compound. A change in the activity of LIPAM in the presence of the test compound is indicative of a compound that modulates the activity of LIPAM. Alternatively, a test compound is combined with an in vitro or cell-free system comprising LIPAM under conditions suitable for LIPAM activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of LIPAM may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding LIPAM or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding LIPAM may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding LIPAM can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region

of a polynucleotide encoding LIPAM is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress LIPAM, e.g., by secreting LIPAM in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

### THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of LIPAM and lipid-associated molecules. In addition, examples of tissues expressing LIPAM are normal lung, cancerous lung, and diseased thyroid tissue, and also can be found in Table 6. Therefore, LIPAM appears to play a role in cancers, neurological, autoimmune/inflammatory, gastrointestinal, and cardiovascular disorders, and disorders of lipid metabolism. In the treatment of disorders associated with increased LIPAM expression or activity, it is desirable to decrease the expression or activity of LIPAM. In the treatment of disorders associated with decreased LIPAM expression or activity, it is desirable to increase the expression or activity of LIPAM.

Therefore, in one embodiment, LIPAM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LIPAM. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive

pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity

5 pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; a neurological disorder such

10 as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial

15 thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral

20 palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive

25 dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune

30 polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis,

hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, 5 fungal, parasitic, protozoal, and helminthic infections, and trauma; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the 10 intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome 15 (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular 20 hyperplasias, adenomas, and carcinomas; and a disorder of lipid metabolism such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier 25 disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay- 30 Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity.

In another embodiment, a vector capable of expressing LIPAM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LIPAM including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified LIPAM in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LIPAM including, but not limited to, those provided above.

- 5 In still another embodiment, an agonist which modulates the activity of LIPAM may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of LIPAM including, but not limited to, those listed above.

In a further embodiment, an antagonist of LIPAM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of LIPAM. Examples of such  
10 disorders include, but are not limited to, those cancers, neurological, autoimmune/inflammatory, gastrointestinal, and cardiovascular disorders, and disorders of lipid metabolism, described above. In one aspect, an antibody which specifically binds LIPAM may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express LIPAM.

- 15 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding LIPAM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of LIPAM including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate  
20 therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

- 25 An antagonist of LIPAM may be produced using methods which are generally known in the art. In particular, purified LIPAM may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind LIPAM. Antibodies to LIPAM may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and  
30 fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J.

Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with LIPAM or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to LIPAM have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of LIPAM amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to LIPAM may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce LIPAM-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter,



G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for LIPAM may also be generated. For example, such fragments include, but are not limited to,  $F(ab)_2$  fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the  $F(ab)_2$  fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between LIPAM and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering LIPAM epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for LIPAM. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of LIPAM-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple LIPAM epitopes, represents the average affinity, or avidity, of the antibodies for LIPAM. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular LIPAM epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the LIPAM-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of LIPAM, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of LIPAM-antibody

complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding LIPAM, or any  
5 fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding LIPAM. Such technology is well known in the art, and antisense oligonucleotides or larger  
10 fragments can be designed from various locations along the coding or control regions of sequences encoding LIPAM. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence  
15 complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other  
20 gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding LIPAM may be used for  
25 somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475),  
30 cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii)

express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in LIPAM expression or regulation causes disease, the expression of LIPAM from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

10 In a further embodiment of the invention, diseases or disorders caused by deficiencies in LIPAM are treated by constructing mammalian expression vectors encoding LIPAM and introducing these vectors by mechanical means into LIPAM-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and  
15 (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of LIPAM include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors  
20 (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). LIPAM may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci.  
25 USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous  
30 gene encoding LIPAM from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental

parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

- 5 In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to LIPAM expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding LIPAM under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences
- 10 required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al.
- 15 (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference.
- 20 Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).
- 25 In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding LIPAM to cells which have one or more genetic abnormalities with respect to the expression of LIPAM. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas
- 30 (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both

incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding LIPAM to target cells which have one or more genetic abnormalities with respect to the expression of LIPAM. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing LIPAM to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding LIPAM to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for LIPAM into the alphavirus genome in place of the capsid-coding region results in the production of a large number of LIPAM-coding RNAs and the synthesis of high levels of LIPAM in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy

application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of LIPAM into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding LIPAM.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding LIPAM. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA

constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

10 An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding LIPAM. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular  
15 chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased LIPAM expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding LIPAM may be therapeutically useful, and in the treatment of disorders associated with  
20 decreased LIPAM expression or activity, a compound which specifically promotes expression of the polynucleotide encoding LIPAM may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in  
25 altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding LIPAM is exposed to at least one test compound thus obtained. The sample  
30 may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding LIPAM are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence

of the polynucleotide encoding LIPAM. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the

5 polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library

10 of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells

15 taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of

20 such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various

25 formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of LIPAM, antibodies to LIPAM, and mimetics, agonists, antagonists, or inhibitors of LIPAM.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal,

30 intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the



case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising LIPAM or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, LIPAM or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example LIPAM or fragments thereof, antibodies of LIPAM, and agonists, antagonists or inhibitors of LIPAM, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

#### DIAGNOSTICS

In another embodiment, antibodies which specifically bind LIPAM may be used for the diagnosis of disorders characterized by expression of LIPAM, or in assays to monitor patients being treated with LIPAM or agonists, antagonists, or inhibitors of LIPAM. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for LIPAM include methods which utilize the antibody and a label to detect LIPAM in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring LIPAM, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of LIPAM expression. Normal or standard values for LIPAM expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to LIPAM under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of LIPAM expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding LIPAM may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect

and quantify gene expression in biopsied tissues in which expression of LIPAM may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of LIPAM, and to monitor regulation of LIPAM levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding LIPAM or closely related molecules may be used to identify nucleic acid sequences which encode LIPAM. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding LIPAM, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the LIPAM encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:10-18 or from genomic sequences including promoters, enhancers, and introns of the LIPAM gene.

Means for producing specific hybridization probes for DNAs encoding LIPAM include the cloning of polynucleotide sequences encoding LIPAM or LIPAM derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding LIPAM may be used for the diagnosis of disorders associated with expression of LIPAM. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a cardiovascular disorder such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification,

mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion  
5 and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative  
10 interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung  
15 transplantation; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural  
20 empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central  
25 nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD),  
30 akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies,

- ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins,
- 5 erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis,
- 10 thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a gastrointestinal disorder such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis,
- 15 gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation,
- 20 gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha<sub>1</sub>-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of
- 25 pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; and a disorder of lipid metabolism such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid
- 30 lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism,

renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity. The polynucleotide sequences encoding LIPAM may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered LIPAM expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding LIPAM may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding LIPAM may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding LIPAM in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of LIPAM, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding LIPAM, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development

of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

- 5 Additional diagnostic uses for oligonucleotides designed from the sequences encoding LIPAM may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding LIPAM, or a fragment of a polynucleotide complementary to the polynucleotide encoding LIPAM, and will be employed under optimized conditions for identification of a specific gene or
- 10 condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

- In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding LIPAM may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease
- 15 in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding LIPAM are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the
- 20 secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual
- 25 overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).
- 30 SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding

lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations. (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641.)

Methods which may also be used to quantify the expression of LIPAM include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, LIPAM, fragments of LIPAM, or antibodies specific for LIPAM may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to



generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be  
5 quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global  
10 pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is  
15 achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot  
20 is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass  
25 spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for LIPAM to quantify  
30 the levels of LIPAM expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by

a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and  
5 should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid  
10 degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample.  
15 A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological  
20 sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

25 Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of  
30 microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding LIPAM may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence.

Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific  
5 region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop  
10 genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) *Proc. Natl. Acad. Sci. USA* 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic  
15 map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding LIPAM on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as  
20 linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely  
25 localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) *Nature* 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

30 In another embodiment of the invention, LIPAM, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes

between LIPAM and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with LIPAM, or fragments thereof, and washed. Bound LIPAM is then detected by methods well known in the art. Purified LIPAM can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

10 In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding LIPAM specifically compete with a test compound for binding LIPAM. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with LIPAM.

In additional embodiments, the nucleotide sequences which encode LIPAM may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, 20 to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, including U.S. Ser. No. 60/266,910, U.S. Ser. No. 60/276,891, U.S. Ser. No. 60/279,760, U.S. Ser. No. 60/283,818, U.S. Ser. No. 60/276,855, and U.S. Ser. No. 60/285,405, are expressly incorporated 25 by reference herein.

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with

chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## 25 II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a

high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner

5 (Labsystems Oy, Helsinki, Finland).

### III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler  
10 (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides  
15 were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the  
20 techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public  
25 databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic  
30 approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched

sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:10-18. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

#### **IV. Identification and Editing of Coding Sequences from Genomic DNA**

Putative lipid-associated molecules were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpr1 and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to



form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode lipid-associated molecules, the encoded polypeptides were

5 analyzed by querying against PFAM models for lipid-associated molecules. Potential lipid-associated molecules were also identified by homology to Incyte cDNA sequences that had been annotated as lipid-associated molecules. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpr public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the

10 sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or

15 public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### **V. Assembly of Genomic Sequence Data with cDNA Sequence Data**

##### **"Stitched" Sequences**

20 Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information,

25 generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated

30 but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or

genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpi public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended  
5 with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

#### **"Stretched" Sequences**

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases  
10 using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The  
15 GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### **VI. Chromosomal Mapping of LIPAM Encoding Polynucleotides**

20 The sequences which were used to assemble SEQ ID NO:10-18 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:10-18 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available  
25 from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map  
30 position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances

are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease

5 genes map within or in proximity to the intervals indicated above.

## VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel  
10 (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar.

15 The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum} \{ \text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2}) \}}$$

20 The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair  
25 (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced  
30 either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding LIPAM are analyzed with respect to the

tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding LIPAM. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

#### **VIII. Extension of LIPAM Encoding Polynucleotides**

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C,

3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### **IX. Identification of Single Nucleotide Polymorphisms in LIPAM Encoding**

### Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:10-18 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the

5 identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files

10 in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins

15 or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The

20 African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed

25 no allelic variance in this population were not further tested in the other three populations.

### X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:10-18 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide

30 fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a

SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech).

An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

- 5           The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and  
10 compared.

#### **XI. Microarrays**

- The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, *supra*.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned  
15 technologies should be uniform and solid with a non-porous surface (Skena (1999), *supra*). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and  
20 may contain any appropriate number of elements. (See, e.g., Skena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

- Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be  
25 selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser  
30 desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

### Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).



Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

## 5 **Hybridization**

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly  
10 larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

## **Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an  
15 Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-  
20 scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate  
25 filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a  
30 cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore,

are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

For example, component 1824717\_HGG4 of SEQ ID NO:15 showed differential expression in tissue affected by cancer versus normal tissue, as determined by microarray analysis. Matched samples of normal lung tissue and lung tissue affected by squamous cell carcinoma, and matched samples of normal lung tissue and lung tissue affected by adenocarcinoma, were provided by the Roy Castle International Center for Lung Cancer Research (Liverpool, UK). The expression of component 1824717\_HGG4 was altered in lung tissue affected by squamous cell carcinoma and in lung tissue affected by adenocarcinoma. Therefore, SEQ ID NO:15 is useful in diagnostic assays for cancer.

## **XII. Complementary Polynucleotides**

Sequences complementary to the LIPAM-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring LIPAM. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of LIPAM. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the LIPAM-encoding transcript.

## **XIII. Expression of LIPAM**

Expression and purification of LIPAM is achieved using bacterial or virus-based expression

systems. For expression of LIPAM in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express LIPAM upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of LIPAM in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding LIPAM by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, LIPAM is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from LIPAM at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified LIPAM obtained by these methods can be used directly in the assays shown in Examples XVII and XVIII, where applicable.

#### XIV. Functional Assays

LIPAM function is assessed by expressing the sequences encoding LIPAM at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which

contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of LIPAM on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding LIPAM and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding LIPAM and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### **XV. Production of LIPAM Specific Antibodies**

LIPAM substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the LIPAM amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well

described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-LIPAM activity by, for example, binding the peptide or LIPAM to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### 10 XVI. Purification of Naturally Occurring LIPAM Using Specific Antibodies

Naturally occurring or recombinant LIPAM is substantially purified by immunoaffinity chromatography using antibodies specific for LIPAM. An immunoaffinity column is constructed by covalently coupling anti-LIPAM antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing LIPAM are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of LIPAM (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/LIPAM binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and LIPAM is collected.

#### XVII. Identification of Molecules Which Interact with LIPAM

LIPAM, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled LIPAM, washed, and any wells with labeled LIPAM complex are assayed. Data obtained using different concentrations of LIPAM are used to calculate values for the number, affinity, and association of LIPAM with the candidate molecules.

Alternatively, molecules interacting with LIPAM are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

LIPAM may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S.

Patent No. 6,057,101).

### **XVIII. Demonstration of LIPAM Activity**

Selected candidate lipid molecules, such as C4 sterols, oxysterol, apolipoprotein E, and phospholipids, are arrayed in the wells of a multi-well plate. LIPAM, or biologically active fragments thereof, are labeled with  $^{125}\text{I}$  Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) The selected candidate lipid molecules are incubated with the labeled LIPAM and washed. Any wells with labeled LIPAM complex are assayed. Data obtained using different concentrations of LIPAM are used to calculate values for the number, affinity, and association of LIPAM with the candidate molecules. Significant binding of LIPAM to the candidate lipid molecules is indicative of LIPAM activity.

In the alternative, LIPAM activity is determined in a continuous fluorescent transfer assay using as substrate 1-palmitoyl-2-pyrenyldecanoyl-phosphatidylinositol (Phy(10)PI). The assay measures the increase of pyrene monomer fluorescence intensity as a result of the transfer of pyrenylacyl (Pyr(x))-labeled phospholipid from quenched donor vesicles to unquenched acceptor vesicles (Van Paridon et al. (1988) Biochemistry 27:6208-6214). Donor vesicles consist of Pyr(x) phosphatidylinositol (Pyr(x)PI), 2,4,6-trinitrophenylphosphatidylethanolamine (TNP-PE) and egg phosphatidylcholine (PC) in a mol % ratio of 10:10:80 (2 nmol of total phospholipid). Acceptor vesicles consist of phosphatidic acid (PA) and egg PC in a mol % ratio of 5:95 (25-fold excess of total phospholipid). The reaction is carried out in 2 ml of 20 mM Tris-HCl, 5 mM EDTA, 200 mM NaCl (pH 7.4) containing 0.1 mg of BSA at 37°C. The reaction is initiated by the addition of 10-50  $\mu\text{l}$  of LIPAM. Measurements are performed using a fluorimeter equipped with a thermostated cuvette holder and a stirring device. The initial slope of the progress curve is taken as an arbitrary unit of transfer activity (van Tiel, C.M. et al. (2000) J. Biol. Chem. 275:21532-21538; Westerman, J. et al. (1995) J. Biol. Chem. 270:14263-14266).

In the alternative, LIPAM activity is determined by measuring the rate of incorporation of a radioactive fatty acid precursor into fatty acyl-CoA. The final reaction contains 200 mM Tris-HCl, pH 7.5, 2.5 mM ATP, 8 mM  $\text{MgCl}_2$ , 2mM EDTA, 20 mM NaF, 0.1% Triton X-100, 10 mM  $[^3\text{H}]$ oleate,  $[^3\text{H}]$ myristate or  $[^{14}\text{C}]$ decanoate, 0.5 mM coenzyme A, and LIPAM in a total volume of 0.5 ml. The reaction is initiated with the addition of coenzyme A, incubated at 35 °C for 10 min, and terminated by the addition of 2.5 ml of isopropyl alcohol, n-heptane, 1 M  $\text{H}_2\text{SO}_4$  (40:10:1). Radioactive fatty acid is removed by organic extraction using n-heptane. Fatty acyl-CoA formed during the reaction remains in the aqueous fraction and is quantified by scintillation counting (Black, P.N. et al. (1997) J. Biol. Chem. 272: 4896-4904).

In the alternative, LIPAM activity is determined by measuring the degradation of the sphingolipid glucosylceramide. 25-50 microunits glucocerebrosidase are incubated with varying concentrations of LIPAM in a 40  $\mu$ l reaction at 37 °C for 20 min. The final reaction contains 50mM sodium citrate pH 4.5, 20 ng human serum albumin, and 3.125 mM lipids in the form of liposomes, which contain lipids in the following proportions: [ $^{14}$ C]glucosylceramide (3 mol %, 2.4 Ci/mol), cholesterol (23 mol %), phosphatidic acid (20 mol %), phosphatidylcholine (54 mol %). The reaction is stopped by the addition of 160  $\mu$ l chloroform/methanol (2:1) and 20  $\mu$ l 0.1% glucose, and shaking. After centrifugation at 4000 rpm, enzymatically released [ $^{14}$ C]glucose in the aqueous phase is measured in a scintillation counter. LIPAM activity is determined by its effect on increasing the rate of glucosylceramide hydrolysis by glucocerebrosidase (Wilkening, G. et al. J. Biol. Chem. (1998) 273:30271-30278).

In the alternative, LIPAM activity can be demonstrated by an *in vitro* hydrolysis assay with vesicles containing 1-palmitoyl-2-[1- $^{14}$ C]oleoyl phosphatidylcholine (Sigma-Aldrich). LIPAM triglyceride lipase activity and phospholipase A<sub>2</sub> activity are demonstrated by analysis of the cleavage products isolated from the hydrolysis reaction mixture.

Vesicles containing 1-palmitoyl-2-[1- $^{14}$ C]oleoyl phosphatidylcholine (Amersham Pharmacia Biotech.) are prepared by mixing 2.0  $\mu$ Ci of the radiolabeled phospholipid with 12.5 mg of unlabeled 1-palmitoyl-2-oleoyl phosphatidylcholine and drying the mixture under N<sub>2</sub>. 2.5 ml of 150 mM Tris-HCl, pH 7.5, is added, and the mixture is sonicated and centrifuged. The supernatant may be stored at 4 °C. The final reaction mixtures contain 0.25 ml of Hanks buffered salt solution supplemented with 2.0 mM taurochenodeoxycholate, 1.0% bovine serum albumin, 1.0 mM CaCl<sub>2</sub>, pH 7.4, 150  $\mu$ g of 1-palmitoyl-2-[1- $^{14}$ C]oleoyl phosphatidylcholine vesicles, and various amounts of LIPAM diluted in PBS. After incubation for 30 min at 37 °C, 20  $\mu$ g each of lyso-phosphatidylcholine and oleic acid are added as carriers and each sample is extracted for total lipids. The lipids are separated by thin layer chromatography using a two solvent system of chloroform:methanol:acetic acid:water (65:35:8:4) until the solvent front is halfway up the plate. The process is then continued with hexane:ether:acetic acid (86:16:1) until the solvent front is at the top of the plate. The lipid-containing areas are visualized with I<sub>2</sub> vapor; the spots are scraped, and their radioactivity is determined by scintillation counting. The amount of radioactivity released as fatty acids will increase as a greater amount of LIPAM is added to the assay mixture while the amount of radioactivity released as lysophosphatidylcholine will remain low. This demonstrates that LIPAM cleaves at the *sn*-2 and not the *sn*-1 position, as is characteristic of phospholipase A<sub>2</sub> activity.

In the alternative, phospholipase activity of LIPAM is measured by the hydrolysis of a fatty

acyl residue at the sn-1 position of phosphatidylserine. LIPAM is combined with the tritium [<sup>3</sup>H] labeled substrate phosphatidylserine at stoichiometric quantities in a suitable buffer. Following an appropriate incubation time, the hydrolyzed reaction products are separated from the substrates by chromatographic methods. The amount of acylglycerophosphoserine produced is measured by counting tritiated product with the help of a scintillation counter. Various control groups are set up to account for background noise and unincorporated substrate. The final counts represent the tritiated enzyme product [<sup>3</sup>H]-acylglycerophosphoserine, which is directly proportional to the activity of LIPAM in biological samples.

Lipoxygenase activity of LIPAM can be measured by chromatographic methods. Extracted LIPAM lipoxygenase protein is incubated with 100  $\mu$ M [1-<sup>14</sup>C] arachidonic acid or other unlabeled fatty acids at 37°C for 30 min. After the incubation, stop solution (acetonitrile:methanol:water, 350:150:1) is added. The samples are extracted and analyzed by reverse-phase HPLC using a solvent system of methanol/water/acetic acid, 85:15:0.01 (vol/vol) at a flow rate of 1 ml/min. The effluent is monitored at 235 nm and analyzed for the presence of the major arachidonic metabolite such as 12-HPETE (catalyzed by 12-LOX). The fractions are also subjected to liquid scintillation counting. The final counts represent the products, which is directly proportional to the activity of LIPAM in biological samples. For stereochemical analysis, the metabolites of arachidonic acid are analyzed further by chiral phase-HPLC and by mass spectrometry (Sun, D. et al. (1998) J. Biol. Chem. 273:33540-33547).

Sialidase activity of LIPAM is assayed using various substrates, including but not limited to 2'-(4-methylumbelliferyl) $\alpha$ -D-N-acetylneuramic acid, 2'-O-(o-nitrophenyl) $\alpha$ -D-N-acetylneuramic acid, 2'-O-(p-nitrophenyl) $\alpha$ -D-N-acetylneuramic acid, and  $\alpha$ (2-3)- and  $\alpha$ (2-6)-sialyllactose. The reaction mixture contains 30 nmol substrate, 0.2 mg bovine serum albumin, 10  $\mu$ mol sodium acetate (pH 4.6), 0.2 mg Triton X-100, and purified LIPAM (or a sample containing LIPAM). Following incubation at 37° C for 10-30 min, the released sialic acid is quantified using the thiobarbituric acid method (Aminoff, D. (1961) Biochem. J. 81:384-392 ). One unit of sialidase activity is defined as the amount of LIPAM that catalyzes the release of 1 nmol of sialic acid from substrate per hour (Hasegawa, T. et al. (2000) J. Biol. Chem. 275:8007-8015).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious



to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
7472774	1	7472774CD1	10	7472774CB1
2884821	2	2884821CD1	11	2884821CB1
72852842	3	72852842CD1	12	72852842CB1
7484271	4	7484271CD1	13	7484271CB1
7474074	5	7474074CD1	14	7474074CB1
72024970	6	72024970CD1	15	72024970CB1
6131380	7	6131380CD1	16	6131380CB1
643681	8	643681CD1	17	643681CB1
6897474	9	6897474CD1	18	6897474CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	7472774CD1	g4886978	1.6E-161	[Homo sapiens] cytosolic phospholipase A2 beta; cPLA2beta (Song, C. et al. (1999) J. Biol. Chem. 274:17063-17067)
2	2884821CD1	g14669826	0.0	lipoic acid synthase [Mus musculus] (Morikawa, T. et al. (2001) FEBS Lett. 498:16-21)
3	72852842CD1	g4894788	2.5E-126	[Mus musculus] phospholipase C delta-1 (Lee, W.K. et al. (1999) Biochem. Biophys. Res. Commun. 261:393-399)
4	7484271CD1	g2138183	2.7E-11	[Mus musculus] polycystic kidney disease 1 protein (Lohning, C. et al. (1997) Mann. Genome 8:307-311)
5	7474074CD1	g4090960	3.1E-63	[Homo sapiens] phosphatidylserine-specific phospholipase A1 (Nagai, Y. et al. (1999) J. Biol. Chem. 274:11053-11059)
6	72024970CD1	g13560884 g6705987	1.0E-109 1.1E-164	lacrima lipase [Oryctolagus cuniculus] [Mus musculus] phospholipase C-L2 (Otsuki, M. (1999) Biochem. Biophys. Res. Commun. 266:97-103)
7	6131380CD1	g5771350	0.0	[Mus musculus] M-RdgB2 retinal degeneration protein B subtype 2 (Lu, C. et al. (1999) J. Neurosci. 19:7317- 7325)
8	643681CD1	g2618983 g8452870	0.0 1.2E-36	[Mus musculus] membrane-associated phosphatidylinositol transfer protein (Aikawa, Y. et al. (1997) Biochem. Biophys. Res. Commun. 236:559-564)
9	6897474CD1	g6651241	3.8E-188	[Homo sapiens] lipopolysaccharide specific response-68 protein [Mus musculus] TAGL-beta

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7472774CD1	996	S34 S46 S64 S133 S151 S169 S183 S219 S273 S294 S418 S557 S652 S662 S769 S928 T17 T24 T87 T104 T227 T248 T368 T603 T722 T775 T808 T916 T952 Y38 Y664	N201 N362 N718 N834 N914	Transmembrane domains: E276-L301, S696-S720 N terminus is cytosolic  CYTOSOLIC PHOSPHOLIPASE A2 CPLA2 INCLUDES: PHOSPHATIDYLCHOLINE 2 ACYLHYDROLASE LYOPHOSPHOLIPASE HYDROLASE LIPID PD014471: G542-L711, G812-N914	TMAP   BLAST- PRODOR
2	2884821CD1	372	S2 S30 S57 S145 S229 S258 S352 T58 T96 T104 T148 T163 T178 T240 T292 T313 T369		SYNTHETASE LIPOIC ACID LIPSYN LIPOATE IRON-SULFUR SYNTHASE PRECURSOR MITOCHONDRION TRANSIT: PD149846: L80-D135 PD005028: Q311-E357  do SYNTHETASE; LIPOIC; ACID; BIOSYNTHESIS: DM02726 P32875 79-413: K36-K370 DM02726 E36953 1-310: L74-A355 DM02726 G64043 7-320: L74-A355 DM02726 P25845 7-320: N65-A355  Cell attachment sequence: R217-D219	BLAST- PRODOR   BLAST-DOMO
3	72852842CD1	649	S313 S419 S429 S542 S574 T24 T56 T68 T79 T220 T267 T303	N417 N578	C2 domain: L525-T613 Phosphatidylinositol-specific phospholipase C, Xdomain: D156-K300	MOTIFS HMME-PFAM HMME-PFAM

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3			T381 T397 T440 T452 T526		Phosphatidylinositol-specific phospholipase C, Y domain: A389-R506 Phosphatidylinositol-specific phospholipase signature BL50007: L161-G206, T220-Q257, L284-K300, H439-G480, Q600-L636 Phospholipase C signature PR00390: P160-Q178, W186-G206, T283-K300, I444-W465, W465-M483, L614-R624 PHOSPHOLIPASE C PHOSPHODIESTERASE HYDROLASE 1PHOSPHATIDYLINOSITOL4 5BISPHOSPHATE LIPID DEGRADATION TRANSDUCER PHOSPHOINOSITIDESPECIFIC: PD001214: D156-K300 PD001202: L390-R506 PHOSPHOLIPASE 1PHOSPHATIDYLINOSITOL4 5BISPHOSPHATE PHOSPHODIESTERASE HYDROLASE LIPID DEGRADATION TRANSDUCER C CALCIUMBINDING: PD004439: R4-Q155 1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE D: DM00855 P51178 64-472: W5-E332 DM00712 P51178 474-754: K378-V645 DM00855 A48047 58-521: N47-S350 DM00855 P10894 62-503: N47-D328	HMMER-PFAM  BLIMPS-BLOCKS  BLIMPS-PRINTS  BLAST-PRODROM  BLAST-PRODROM  BLAST-DOMO
4	748427LCD1	2020	S76 S163 S245 S436 S578 S598 S693 S705 S706 S710 S711 S712 S809 S812	N341 N1128 N1248 N1296 N1336 N1533 N1553 N1680	PLAT/LH2 (Polycystin-1, Lipoxigenase, Alpha-Toxin/lipoxygenase homology) domain: N769-E885, T1632-F1749, F431-Y550, T922-L1039, V1901-L2016, I43-Y159, V1207-R1322, A300-L419, V1505-C1620, V561-E677, F172-M286, V1053-L1177, I1374-R1489, T1763-E1883	HMMER-PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4			S960 S965 S1139 S1167 S1192 S1312 S1339 S1349 S1406 S1424 S1535 S1578 S1607 S1715 S1723 S1737 S1739 S1800 T16 T292 T360 T427 T462 T656 T927 T949 T963 T996 T1092 T1166 T1173 T1184 T1250 T1298 T1382 T1487 T1745 T1763 T1810 T1852 T1854 T1909 T1932 T1950 T2012 Y771 Y790 Y1180 Y1329 Y1429 Y1975	N1790 N1799 N1988	Transmembrane domains: L31-D53, A1071-R1095, T1487-W1512 N-terminus is non-cytosolic PROTEIN POLYCYSTIC KIDNEY DISEASE REPEAT TRANSMEMBRANE POLYCYSTIN PRECURSOR AUTOSOMAL DOMINANT: PD010179: Y1634-K1793 Cell attachment sequence: R1482-D1484, R2007-D2009 ATP/GTP-binding site motif A (P-loop): G1239-S1246	TMAP  BLAST- PRODOM  MOTIFS  MOTIFS
5	7474074CD1	415	S145 S151 S232 S311 T27 T84 T135 T304 T371 T411	N18 N351	signal cleavage: M1-S43 Lipase: M1-L275 Transmembrane domains: D106-I134, T278-M306 N terminus is non-cytosolic	SPSCAN HMME-PFAM TMAP

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5					<p>Lipases, serine proteins BL00120: N62-I76, D106-S120, Y183-C193</p> <p>Triacylglycerol lipase family signature PR00821: N107-K125, C206-T221, N19-Y38, I64-R79</p> <p>Vespid venom allergen phospholipase A1 signature PR00825: P148-H165, L171-P191</p> <p>LIPASE PRECURSOR SIGNAL HYDROLASE LIPID DEGRADATION GLYCOPROTEIN PANCREATIC PROTEIN PANCREAS: PD001492: N6-L314</p> <p>TRIACYLGLYCEROL LIPASE: DM00344 A49488 25-326: M1-F294 DM00344 P11150 38-356: M1-L296 DM00344 S15893 37-357: F25-F294 DM00344 P11153 17-335: N22-M270</p>	<p>BLIMPS-BLOCKS</p> <p>BLIMPS-PRINTS</p> <p>BLIMPS-PRINTS</p> <p>BLAST-PRODROM</p> <p>BLAST-DOMO</p>
6	72024970CD1	1152	<p>S79 S91 S124</p> <p>S154 S186</p> <p>S235 S276</p> <p>S350 S445</p> <p>S483 S487</p> <p>S543 S550</p> <p>S558 S565</p> <p>S577 S584</p> <p>S591 S649</p> <p>S681 S881</p> <p>S918 S932</p> <p>S980 S1100</p> <p>S1111 T134</p> <p>T173 T236</p> <p>T387 T504</p> <p>T512 T615</p>	<p>N290 N303</p> <p>N472 N534</p>	<p>C2 domain: L756-T848</p> <p>PH domain: A44-A151</p> <p>Phosphatidylinositol-specific phospholipase: D323-K468, A621-C736</p> <p>EF hand: W169-L197, R205-M234</p> <p>Phosphatidylinositol-specific phospholipase X-box domain protein BL50007: F669-G710, D835-I871, L328-G373, T387-Q424, L452-K468</p> <p>C2 domain signature PR00360: R777-I789, N807-M820, V829-D837</p> <p>Phospholipase C signature PR00390: P327-Q345, D353-G373, T451-K468, L674-W695, W695-L713, L849-R859</p>	<p>HMMER-PFAM</p> <p>HMMER-PFAM</p> <p>HMMER-PFAM</p> <p>HMMER-PFAM</p> <p>BLIMPS-BLOCKS</p> <p>BLIMPS-PRINTS</p> <p>BLIMPS-PRINTS</p>

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6			T812 Y628		<p>PHOSPHOLIPASE C PHOSPHODIESTERASE HYDROLASE 1PHOSPHATIDYLINOSITOL4 5BISPHOSPHATE LIPID DEGRADATION TRANSDUCER PHOSPHOINOSITIDE-SPECIFIC: PD001214: D323-K468 PD001202: L622-P732</p> <p>PHOSPHOLIPASE 1PHOSPHATIDYLINOSITOL4 5BISPHOSPHATE PHOSPHODIESTERASE HYDROLASE LIPID DEGRADATION TRANSDUCER C CALCIUM-BINDING: PD004439: Q100-Q322</p> <p>1-PHOSPHATIDYLINOSITOL-4, 5-BIS- PHOSPHATE PHOSPHODIESTERASE D DM00855: P51178 64-472: S88-D494 P08487 71-500: I87-G500 P16885 63-486: I87-E485 P40977 208-616: I89-N499</p> <p>EF-hand calcium-binding domain: D178-V190</p>	BLAST- PRODOM
7	6131380CD1	1294	<p>S29 S38 S129 S164 S205 S254 S290 S307 S310 S313 S314 S324 S340 S341 S353 S367 S399 S443 S444 S496 S504 S584 S585 S589 S592</p>	<p>N31 N652 N911 N945 N1006</p>	<p>Phosphatidylinositol transfer protein: M1-L253 Transmembrane domain: G553-C572, A711-K737, T1050-V1065 N-terminus is non-cytosolic</p> <p>Phosphatidylinositol transfer protein signature PR00391: F198-D213, V219- S238, E16-G35, V85-E105, I111-F126</p> <p>PROTEIN PHOSPHATIDYLINOSITOL TRANSFER ISOFORM PTD INS PTD INSTP LIPID BINDING TRANSPORT ALPHA PIT P ALPHA: PD006368: M1-M257</p>	<p>HMMER-PFAM</p> <p>TMAP</p> <p>BLIMPS- PRINTS</p> <p>BLAST- PRODOM</p>



Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7			S633 S637S644 S666 S669 S767 S833 S891 S913 S1008 S1018 S1121 S1155 S1203 S1222 S1250 S1258 T13 T33 T59 T155 T167 T280 T433 T704 T799 T939 T966 T1014 T1037 T1096 T1179 T1282 Y671 Y871 Y1090		PROTEIN RETINAL DEGENERATION B PHOSPHATIDYLINOSITOL MEMBRANE ASSOCIATED HOMOLOGUE OF DROSOPHILA GENE PD018514: P786-K1048 PD025569: G1049-L1200 PD018515: R396-I565 PHOSPHATIDYLINOSITOL; BETA; DEGENERATION; TRANSFER; DM02192: P43125 1-222: M1-D213 Q00169 1-213: M1-H212 P53812 1-213: M1-I211, A850-D867 JX0316 1-214: M1-I211, A850-D867 Leucine zipper pattern: L1218-L1239 Cell attachment sequence: R1033-D1035	BLAST- PRODROM
8	643681CD1	77	T31 T32 T41		Signal cleavage: M1-P67	SPSCAN
9	6897474CD1	576	S202 S239 S558 S561 T79 T154 T181 T213 T259 T498 T548	N77 N367 N485	Signal cleavage: M1-A21 Signal Peptide: M1-S22 Transmembrane domains: V215-G234, P255-G283 N-terminus is cytosolic PROTEIN PEPTIDOGLYCAN RECOGNITION PRECURSOR SIGNAL TUMOR ASSOCIATED CSP PD090970: A368-Y486	SPSCAN HMNER TMAP BLAST- PRODROM

Table 4

Polynucleotide SEQ ID NO: / Incyte ID / Sequence Length	Sequence Fragments
10/7472774CB1/ 3879	1-326, 1-558, 310-558, 463-2162, 665-811, 1490-2112, 1491-1672, 1491-1787, 1491-1823, 1491-1889, 1491-1892, 1491-1958, 1491-1972, 1491-2022, 1491-2025, 1491-2119, 1491-2120, 1493-2120, 1496-2120, 1497-2050, 1499-2120, 1557-2297, 1587-1985, 1648-2120, 1675-2297, 1767-2120, 1807-2166, 1854-2297, 1857-2297, 1926-2297, 1944-2150, 1944-3300, 2277-2949, 2281-2524, 2281-2898, 2281-2949, 2282-2947, 2325-2934, 2330-2935, 2340-2935, 2347-2934, 2387-2910, 2389-2934, 2395-2949, 2417-2524, 2445-2935, 2464-2934, 2466-2934, 2488-2934, 2503-2935, 2514-2935, 2532-2934, 2563-2935, 2597-2935, 2729-2878, 2729-3192, 2729-3249, 2729-3339, 2729-3399, 2868-3420, 2934-3473, 2988-3592, 3066-3823, 3121-3726, 3179-3879, 3182-3875, 3182-3876, 3182-3878, 3182-3879, 3183-3879, 3184-3879, 3199-3565, 3220-3549, 3228-3802, 3269-3410, 3269-3646, 3291-3879
11/2884821CB1/ 1623	1-318, 52-352, 60-246, 60-318, 60-362, 60-371, 60-743, 61-307, 63-314, 63-341, 81-391, 88-366, 96-371, 99-725, 107-351, 111-803, 124-380, 168-791, 231-672, 259-905, 263-932, 368-636, 381-610, 381-845, 389-658, 389-833, 420-1067, 471-698, 524-1054, 654-1087, 681-1236, 721-1255, 765-1317, 770-1271, 838-1233, 856-1078, 856-1245, 856-1321, 856-1512, 879-1524, 952-1152, 1021-1496, 1050-1623, 1076-1449, 1089-1623, 1136-1313, 1153-1595, 1182-1333
12/72852842CB1/ 2199	1-445, 1-550, 1-853, 336-1614, 1197-1839, 1200-1840, 1249-1569, 1249-1573, 1251-1569, 1251-1573, 1288-1569, 1288-1573, 1300-2058, 1336-2135, 1343-1818, 1383-2137, 1551-1814, 1551-2164, 1577-2137, 1683-1835, 1727-2137, 1734-2197, 1758-2198, 1765-2197, 1768-2137, 1769-2197, 1791-2137, 1798-2137, 1851-2137, 2004-2137, 2004-2199, 2042-2137, 2058-2197
13/7484271CB1/ 6326	1-130, 1-131, 1-341, 1-522, 1-712, 22-131, 35-131, 36-129, 67-679, 69-91, 70-91, 129-149, 129-150, 158-285, 158-288, 158-527, 158-652, 158-672, 179-648, 192-858, 193-684, 226-248, 227-248, 234-1071, 264-1071, 443-1071, 451-1071, 456-1070, 523-1071, 535-1071, 558-1071, 565-1071, 568-1071, 574-1071, 577-1071, 591-1071, 869-1267, 1177-1440, 1177-1699, 1321-2700, 1373-2040, 1374-1821, 1374-2040, 1377-1998, 1377-2040, 1613-1935, 1884-2040, 1889-2040, 1894-2040, 2061-2421, 2061-2434, 2061-2504, 2061-2553, 2061-2570, 2061-2576, 2061-2628, 2061-2671, 2061-2674, 2061-2738, 2061-2812, 2062-2253, 2062-2448, 2062-2572, 2062-2590, 2062-2672, 2062-2673, 2062-2687, 2066-2639, 2129-2891, 2140-2639, 2145-2646, 2149-2674, 2152-2768, 2155-2674, 2156-2783, 2150-2778, 2196-2930, 2198-2918, 2212-2513, 2212-2639, 2212-2643, 2212-2666, 2212-2670, 2212-2671, 2212-2683, 2212-2689, 2212-2700, 2218-2700, 2233-2904, 2234-2952, 2249-2700, 2251-2907, 2251-2912, 2266-2700, 2493-2964, 2494-2946, 2538-2700, 2553-2700, 2608-2700, 2691-3263, 2738-3245, 2741-3238, 2741-3263,

Table 4 (cont.)

Polynucleotide SEQ ID NO: / Incyte ID / Sequence Length 13 (cont.)	Sequence Fragments
	2741-3266, 2741-3273, 2741-3326, 2837-3346, 2837-3362, 2837-3368, 2837-3451, 2865-4748, 2870-3645, 2887-3575, 2893-3405, 2913-3195, 2916-3489, 2921-3648, 2926-3548, 2932-3332, 2933-3642, 2937-3271, 2938-3648, 2943-3648, 2952-3648, 2975-3648, 2977-3342, 2978-3439, 2986-3640, 2989-3648, 2999-3648, 3012-3648, 3031-3648, 3032-3551, 3033-3606, 3037-3648, 3044-3648, 3059-3648, 3063-3641, 3074-3646, 3092-3648, 3103-3648, 3107-3644, 3111-3648, 3117-3648, 3125-3642, 3128-3648, 3132-3648, 3137-3648, 3141-3648, 3145-3648, 3151-3648, 3157-3648, 3158-3648, 3159-3648, 3193-3648, 3211-3648, 3249-3648, 3319-3648, 3479-3648, 3653-3951, 4338-4616, 4338-4865, 4397-4660, 4397-4762, 4397-5019, 4464-5142, 4542-5043, 4633-5264, 4658-5249, 4722-5199, 4723-5268, 4773-5280, 4798-5285, 4867-5400, 4890-5522, 4967-5582, 5024-5613, 5045-5079, 5045-5083, 5045-5084, 5045-5091, 5048-5706, 5158-5713, 5178-5779, 5214-5890, 5240-5889, 5269-5815, 5275-5892, 5290-5929, 5363-5620, 5373-5951, 5381-5924, 5397-5747, 5478-6326, 5539-5915, 5801-6246, 5801-6298, 5801-6326, 5833-6326, 5852-5898, 5855-5898, 5856-5898
14/7474074CB1/ 1561	1-262, 132-1114, 260-322, 646-763, 646-771, 646-897, 718-771, 718-856, 718-857, 718-859, 718-860, 770-1161, 1099-1561, 1100-1290, 1100-1556, 1100-1561, 1165-1561, 1273-1561, 1290-1559, 1302-1559, 1395-1556
15/72024970CB1/ 4941	1-698, 1-722, 1-740, 1-744, 1-753, 1-762, 1-764, 1-818, 5-506, 109-686, 123-681, 160-756, 212-870, 281-804, 299-824, 338-862, 389-615, 418-1240, 426-1240, 428-822, 430-885, 431- 1240, 435-1019, 443-1240, 445-1094, 445-1107, 448-1240, 451-1138, 454-1000, 454-1240, 456-1240, 463-1056, 470-1019, 471-1019, 471-1240, 488-646, 488-1191, 488-1217, 488-1223, 488-1236, 488-1240, 491-1240, 511-1008, 516-958, 520-979, 532-1240, 534-1240, 538-1166, 541-1240, 548-1132, 571-1239, 578-1240, 582-1153, 607-1011, 607-1240, 622-1094, 626-1240, 634-1240, 637-1240, 639-666, 641-686, 645-1240, 654-946, 654-1198, 654-1199, 657-1240, 663-1240, 668-1240, 670-1240, 675-1240, 676-1240, 677-761, 682-1240, 684-1240, 685-1240, 696-1240, 703-1240, 728-1240, 736-1240, 737-1166, 744-1084, 747-1554, 754-1240, 798-1240, 807-1240, 812-1240, 823-1240, 831-1107, 852-1240, 854-1240, 873-1239, 873-1240, 877-1240, 903-1240, 909-1240, 910-1554, 912-1240, 913-1240, 915-1240, 928-1240, 941-1240, 949-1077, 949-1379, 949-1418, 949-1423, 949-1437, 949-1462, 949-1463, 949-1469, 949-1473, 949-1493, 950-1493, 951-1493, 955-1493, 956-1493, 982-1240, 982-1487, 982-1497, 988-1012, 988-1015, 988-1016, 990-1015, 994-1229, 994-1236, 994-1240, 996-1240, 996-1565, 1031-1493, 1090- 1631, 1090-1668, 1090-1670, 1096-1635, 1096-1689, 1100-1630, 1119-1548, 1134-1665, 1134- 1698, 1134-1734, 1157-1740, 1158-1687, 1158-1689, 1165-1771, 1237-1493, 1259-1391,

Table 4 (cont.)

Polynucleotide SEQ ID NO: / Incyte ID / Sequence Length	Sequence Fragments
15 (cont.)	1275-1605, 1276-1493, 1347-1933, 1358-1896, 1369-1844, 1369-1860, 1369-1865, 1369-1896, 1369-1900, 1369-1977, 1374-1495, 1380-1890, 1404-1874, 1499-2026, 1499-2036, 1508-1976, 1524-2192, 1526-2192, 1539-2050, 1554-2192, 1569-2193, 1572-2026, 1593-2192, 1594-2192, 1604-2192, 1605-2191, 1609-2192, 1620-2192, 1624-2192, 1632-2192, 1637-2192, 1652-2191, 1652-2192, 1666-2192, 1671-2192, 1735-1784, 1735-2192, 1747-2192, 1757-2192, 1764-2192, 1767-2192, 1776-2261, 1790-2406, 1839-2192, 1872-2192, 1888-2192, 1895-2192, 1916-2191, 1916-2192, 1922-2192, 1936-2192, 1947-2192, 1998-2617, 2062-2625, 2093-2192, 2094-2193, 2128-2434, 2128-2749, 2132-2291, 2136-2754, 2199-2736, 2238-2980, 2287-2969, 2294-2558, 2303-2885, 2304-2980, 2309-2759, 2309-2960, 2356-3100, 2371-2989, 2456-3077, 2491-2945, 2540-2808, 2610-3047, 2611-3153, 2648-3156, 2668-2880, 2672-3326, 2787-3622, 2810-3135, 2810-3139, 2813-3383, 2814-3244, 2814-3267, 2814-3360, 2814-3376, 2814-3377, 2819-3376, 2823-3321, 2824-3374, 2863-3217, 2883-3218, 2956-3377, 2996-3555, 3048-3803, 3092-3718, 3113-3657, 3120-3215, 3205-3489, 3267-3485, 3355-3718, 3485-3715, 3505-3718, 3539-3739, 3548-3718, 3551-3987, 3566-3716, 3610-3912, 3721-3875, 3744-3799, 3745-4088, 3746-4083, 3759-4011, 3820-4344, 3829-4078, 3849-3997, 3939-4509, 4064-4284, 4152-4632, 4213-4468, 4213-4736, 4213-4761, 4213-4868, 4213-4941, 4227-4386, 4239-4468, 4247-4744, 4262-4543
16/6131380CB1/ 4159	1-353, 54-579, 293-934, 349-981, 351-522, 376-643, 460-708, 488-809, 545-690, 667-1226, 704-1358, 805-1358, 911-1358, 1026-1226, 1030-1717, 1037-1467, 1039-1467, 1044-1598, 1103-1714, 1119-1714, 1133-1717, 1316-1985, 1322-1753, 1362-1691, 1497-1657, 1575-1985, 1715-2268, 1743-1938, 1766-2201, 1771-2095, 1835-2332, 1885-2434, 1975-2329, 1993-2632, 2003-2507, 2054-2991, 2267-2507, 2326-2507, 2330-2507, 2451-2507, 2888-4159
17/643681CB1/ 1481	1-299, 51-328, 51-335, 84-353, 92-342, 95-315, 107-370, 128-376, 139-459, 168-431, 181- 447, 191-421, 196-606, 217-516, 287-520, 335-578, 344-599, 344-880, 361-659, 379-637, 385-617, 458-745, 541-768, 541-805, 544-788, 544-1065, 681-1250, 708-939, 718-1299, 726- 969, 747-1399, 757-965, 772-1377, 794-1077, 796-1399, 811-1057, 846-1394, 864-1120, 865- 1127, 871-1107, 871-1341, 871-1399, 873-1402, 904-1142, 904-1169, 975-1214, 1003-1242, 1021-1286, 1057-1245, 1098-1281, 1112-1365, 1173-1391, 1173-1398, 1173-1427, 1173-1430, 1237-1481
18/6897474CB1/ 1841	1-534, 1-1841, 6-150, 7-196, 43-379, 43-572, 59-584, 228-744, 228-810, 515-1169, 805- 1409, 889-1436, 889-1455, 1216-1766, 1287-1834, 1486-1734, 1526-1836, 1526-1841

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
10	7472774CB1	MYEPTXT02
11	2884821CB1	TYMNOT08
12	72852842CB1	TESTNOT17
13	7484271CB1	BONREFC01
14	7474074CB1	UTRSNOR01
15	72024970CB1	LIVRTXS02
16	6131380CB1	NERDTN03
17	643681CB1	PENITUT01
18	6897474CB1	LIVRTMR01

Table 6

Library	Vector	Library Description
BONRFEC01	pINCY	This large size-fractionated library was constructed using RNA isolated from rib bone tissue removed from a Caucasian male fetus who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation. Serologies were negative.
LIVRTMR01	PCDNA2.1	This random primed library was constructed using RNA isolated from liver tissue removed from a 62-year-old Caucasian female during partial hepatectomy and exploratory laparotomy. Pathology for the matched tumor tissue indicated metastatic intermediate grade neuroendocrine carcinoma, consistent with islet cell tumor, forming nodules ranging in size, in the lateral and medial left liver lobe. The pancreas showed fibrosis, chronic inflammation and fat necrosis consistent with pseudocyst. The gallbladder showed mild chronic cholecystitis. Patient history included malignant neoplasm of the pancreas tail, pulmonary embolism, hyperlipidemia, thrombophlebitis, joint pain in multiple joints, type II diabetes, benign hypertension, cerebrovascular disease, and normal delivery. Previous surgeries included distal pancreatectomy, total splenectomy, and partial hepatectomy. Family history included pancreas cancer with secondary liver cancer, benign hypertension, and hyperlipidemia.
LIVRTXS02	pINCY	This subtracted C3A liver tumor cell line tissue library was constructed using 6.4 million clones from a treated C3A hepatocyte cell line library and was subjected to two rounds of subtraction hybridization with 1.72 million clones from an untreated C3A hepatocyte cell line library. The starting library for subtraction was constructed using RNA isolated from a treated C3A hepatocyte cell line which is a derivative of Hep G2, a cell line derived from a hepatoblastoma removed from a 15-year-old Caucasian male. The cells were treated with 3-methylcholanthrene (MCA). The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from untreated C3A hepatocyte cells tissue from the same cell line. Subtractive hybridization conditions were based on the methodologies of Swaroop, et al., NAR 19 (1991):1954 and Bonaldo, et al. Genome Research 6 (1996):791.
MYEPTXT02	pINCY	The library was constructed using RNA isolated from a treated K-562 cell line, derived from chronic myelogenous leukemia precursor cells removed from a 53-year-old female. The cells were treated with 1 micromolar PMA for 96 hours.
NERDTN03	pINCY	This normalized dorsal root ganglion tissue library was constructed from 1.05 million independent clones from a dorsal root ganglion tissue library. Starting RNA was made from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema, acute bronchopneumonia, bilateral pleural effusions, pericardial effusion, and malignant lymphoma (natural killer cell

Table 6 (cont.)

Library	Vector	Library Description
NERDTDN03 (cont.)		type). The patient presented with pyrexia of unknown origin, malaise, fatigue, and gastrointestinal bleeding. Patient history included probable cytomegalovirus infection, liver congestion, and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, respiratory failure, pneumonia of the left lung, natural killer cell lymphoma of the pharynx, Bell's palsy, and tobacco and alcohol abuse. Previous surgeries included colonoscopy, closed colon biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy. Patient medications included Diflucan (fluconazole), Deltasone (prednisone), hydrocodone, Lortab, Alprazolam, Reazodone, ProMace-Cytabom, Etoposide, Cisplatin, Cytarabine, and dexamethasone. The patient received radiation therapy and multiple blood transfusions. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
PENITUT01	pINCY	Library was constructed using RNA isolated from tumor tissue removed from the penis of a 64-year-old Caucasian male during penile amputation. Pathology indicated a fungating invasive grade 4 squamous cell carcinoma involving the inner wall of the foreskin and extending onto the glans penis. Patient history included benign neoplasm of the large bowel, atherosclerotic coronary artery disease, angina pectoris, gout, and obesity. Family history included malignant pharyngeal neoplasm, chronic lymphocytic leukemia, and chronic liver disease.
TESTNOT17	pINCY	Library was constructed from testis tissue removed from a 26-year-old Caucasian male who died from head trauma due to a motor vehicle accident. Serologies were negative. Patient history included a hernia at birth, tobacco use (1 1/2 ppd), marijuana use, and daily alcohol use (beer and hard liquor).
TYMNOT08	pINCY	The library was constructed using RNA isolated from anergic/allogenic T-lymphocyte tissue removed from an adult (40-50-year-old) Caucasian male. The cells were incubated for 3 days in the presence of 1 microgram/ml OKT3 mAb and 5% human serum.
UTRSNOR01	pINCY	Library was constructed using RNA isolated from uterine endometrium tissue removed from a 29-year-old Caucasian female during a vaginal hysterectomy and cystocele repair. Pathology indicated the endometrium was secretory, and the cervix showed mild chronic cervicitis with focal squamous metaplasia. Pathology for the associated tumor tissue indicated intramural uterine leiomyoma. Patient history included hypothyroidism, pelvic floor relaxation, and paraplegia. Family history included benign hypertension, type II diabetes, and hyperlipidemia.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Score= 0 or greater



Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting  
5 of SEQ ID NO:1-9,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90%  
identical to an amino acid sequence selected from the group consisting of SEQ ID  
NO:1, SEQ ID NO:5-6, and SEQ ID NO:9,
  - c) a polypeptide comprising a naturally occurring amino acid sequence at least 91%  
10 identical to an amino acid sequence selected from the group consisting of SEQ ID  
NO:2-4 and SEQ ID NO:7,
  - d) a polypeptide comprising a naturally occurring amino acid sequence at least 99%  
identical to the amino acid sequence of SEQ ID NO:8,
  - e) a biologically active fragment of a polypeptide having an amino acid sequence  
15 selected from the group consisting of SEQ ID NO:1-9, and
  - f) an immunogenic fragment of a polypeptide having an amino acid sequence selected  
from the group consisting of SEQ ID NO:1-9.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the  
20 group consisting of SEQ ID NO:1-9.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.  
25
5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from  
the group consisting of SEQ ID NO:10-18.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a  
30 polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:
- 5       a)     culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
- b)     recovering the polypeptide so expressed.
- 10       10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.
11. An isolated antibody which specifically binds to a polypeptide of claim 1.
- 15       12. An isolated polynucleotide selected from the group consisting of:
- a)     a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:10-18,
- b)     a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ
- 20       ID NO:10-16 and SEQ ID NO:18,
- c)     a polynucleotide comprising a naturally occurring polynucleotide sequence at least 91% identical to the polynucleotide sequence of SEQ ID NO:17,
- d)     a polynucleotide complementary to a polynucleotide of a),
- e)     a polynucleotide complementary to a polynucleotide of b),
- 25       f)     a polynucleotide complementary to a polynucleotide of c), and
- g)     an RNA equivalent of a)-f).
13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.
- 30       14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:
- a)     hybridizing the sample with a probe comprising at least 20 contiguous nucleotides

comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

- 5        b)     detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

- 10        16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- a)     amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and  
          b)     detecting the presence or absence of said amplified target polynucleotide or fragment  
15           thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

- 20        18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

19. A method for treating a disease or condition associated with decreased expression of functional LIPAM, comprising administering to a patient in need of such treatment the composition of  
25     claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a)     exposing a sample comprising a polypeptide of claim 1 to a compound, and  
30           b)     detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional LIPAM, comprising administering to a patient in need of such treatment a composition of claim 21.

5           23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a)       exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b)       detecting antagonist activity in the sample.

10           24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional LIPAM, comprising administering to a patient in need of such treatment a composition of claim 24.

15

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- a)       combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- 20       b)       detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- 25       a)       combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b)       assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c)       comparing the activity of the polypeptide of claim 1 in the presence of the test
- 30       compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A diagnostic test for a condition or disease associated with the expression of LIPAM in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,

- c) a Fab fragment,
- d) a F(ab')<sub>2</sub> fragment, or
- e) a humanized antibody.

5           32. A composition comprising an antibody of claim 11 and an acceptable excipient.

          33. A method of diagnosing a condition or disease associated with the expression of LIPAM  
in a subject, comprising administering to said subject an effective amount of the composition of claim  
32.

10

          34. A composition of claim 32, wherein the antibody is labeled.

          35. A method of diagnosing a condition or disease associated with the expression of LIPAM  
in a subject, comprising administering to said subject an effective amount of the composition of claim

15   34.

          36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim  
11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence  
20           selected from the group consisting of SEQ ID NO:1-9, or an immunogenic fragment  
          thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal  
antibody which binds specifically to a polypeptide comprising an amino acid sequence  
25           selected from the group consisting of SEQ ID NO:1-9.

          37. A polyclonal antibody produced by a method of claim 36.

          38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

30

          39. A method of making a monoclonal antibody with the specificity of the antibody of claim  
11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence

selected from the group consisting of SEQ ID NO:1-9, or an immunogenic fragment thereof, under conditions to elicit an antibody response,

- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide



comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-9.

5 13. 46. A microarray wherein at least one element of the microarray is a polynucleotide of claim

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- 10 b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

15 48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

20 49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

25 51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

30 53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

10 56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

15

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

20

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

25

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:10.

30

66. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:11.

67. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:12.

68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.

69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.

5 70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

10

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

<110> INCYTE GENOMICS, INC.

DAS, Debopriya  
 YAO, Monique G.  
 ARVIZU, Chandra  
 BAUGHN, Mariah R.  
 LU, Yan  
 HAFALIA, April J.A.  
 WALIA, Narinder K.  
 GRIFFIN, Jennifer A.  
 LU, Dyung Aina M.  
 YUE, Henry  
 DING, Li  
 TOWNLEY, David J.  
 ELLIOTT, Vicki S.  
 FORSYTHE, Ian J.  
 RAMKUMAR, Jayalaxmi  
 GANDHI, Ameena R.  
 ISON, Craig H.  
 WARREN, Bridget A.  
 TANG, Y. Tom  
 EMERLING, Brooke M.  
 HONCHELL, Cynthia D.

<120> LIPID-ASSOCIATED MOLECULES

<130> PI-0358 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/266,910; 60/276,891; 60/276,855; 60/279,760;  
 60/283,818; 60/285,405

<151> 2001-02-06; 2001-03-16; 2001-03-16; 2001-03-28;  
 2001-04-13; 2001-04-20

<160> 18

<170> PERL Program

<210> 1

<211> 996

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 7472774CD1

<400> 1

Met	Lys	Arg	Ser	Arg	Pro	Met	His	Pro	Ile	Cys	Leu	Pro	Thr	Gln
1				5					10					15
Thr	Thr	Pro	Arg	Ala	Ile	Pro	Ala	Thr	Ala	Lys	Leu	Trp	Pro	Gly
			20						25					30
Arg	Trp	Ser	Ser	Glu	Ser	Glu	Tyr	Lys	Phe	Leu	Ile	Leu	Pro	Pro
			35						40					45
Ser	Trp	Arg	Ala	Ala	Val	Met	Leu	Leu	Arg	Gln	Met	His	Ala	Arg

	50		55		60
Val Ser His Ser Leu Pro Asp Pro Cys Gln Ala Glu Asp Ser Arg					
	65		70		75
Pro Ser Ala Thr Cys Ala Leu Lys Ala Pro Gln Thr Ser Trp Asp					
	80		85		90
Gly Leu Leu Arg Glu Gly Leu Ser Pro Cys His Leu Leu Thr Val					
	95		100		105
Arg Val Ile Arg Met Lys Asn Val Arg Gln Ala Asp Met Gln Pro					
	110		115		120
Val Gly Ile Glu Leu Ala Pro Cys Leu Gln Ala Pro Ser Val Pro					
	125		130		135
Glu Thr Asp Leu Lys Gly Val Val Gln Ala Arg Gly Gly Gly Ala					
	140		145		150
Ser Val Leu Glu Lys Pro Arg Glu Gly Phe Lys Arg Ala Glu Gln					
	155		160		165
Val Pro Val Ser Gln Thr Asp Cys Phe Val Ser Leu Trp Leu Pro					
	170		175		180
Thr Ala Ser Gln Lys Lys Leu Arg Thr Arg Thr Ile Ser Asn Cys					
	185		190		195
Pro Asn Pro Glu Trp Asn Glu Ser Phe Asn Phe Gln Ile Gln Ser					
	200		205		210
Arg Val Lys Asn Val Leu Glu Leu Ser Val Cys Asp Glu Asp Thr					
	215		220		225
Val Thr Pro Asp Asp His Leu Leu Thr Val Leu Tyr Asp Leu Thr					
	230		235		240
Lys Leu Cys Phe Arg Lys Lys Thr His Val Lys Phe Pro Leu Asn					
	245		250		255
Pro Gln Gly Met Glu Glu Leu Glu Val Glu Phe Leu Leu Glu Glu					
	260		265		270
Ser Pro Ser Pro Pro Glu Thr Leu Val Thr Asn Gly Val Leu Val					
	275		280		285
Val Ile Ile Phe Leu Gly Ser Cys Ser Ser Arg Gly His Gly Trp					
	290		295		300
Leu Leu Leu Ser Gly Glu Gln Asp Gln Gly Arg Lys Gln Trp Ala					
	305		310		315
Gln Leu Gly Leu Cys Pro Ile Leu Thr Ser Ala Gly Val Arg Leu					
	320		325		330
Asn Glu Ala Ser Gln Met Gly His Arg Gln His Trp Gly Thr Ser					
	335		340		345
Trp Gly Phe Cys Thr Glu Gly Gly Val Lys Asp Leu Leu Val Met					
	350		355		360
Val Asn Glu Ser Phe Glu Asn Thr Gln Arg Val Arg Pro Cys Leu					
	365		370		375
Glu Pro Cys Cys Pro Thr Ser Ala Cys Phe Gln Thr Ala Ala Cys					
	380		385		390
Phe His Tyr Pro Lys Tyr Phe Gln Ser Gln Val His Val Glu Val					
	395		400		405
Pro Lys Ser His Trp Ser Cys Gly Leu Cys Cys Arg Ser Arg Lys					
	410		415		420
Lys Gly Pro Ile Ser Gln Pro Leu Asp Cys Leu Ser Asp Gly Gln					
	425		430		435
Val Met Thr Leu Pro Val Gly Glu Ser Tyr Glu Leu His Met Lys					
	440		445		450
Ser Thr Pro Cys Pro Glu Thr Leu Asp Val Arg Leu Gly Phe Ser					
	455		460		465
Leu Cys Pro Ala Glu Leu Glu Phe Leu Gln Lys Arg Lys Val Val					

	470		475		480
Val Ala Lys Ala	Leu Lys Gln Val Leu	Gln Leu Glu Glu Asp	Leu		
	485		490		495
Gln Glu Asp Glu	Val Pro Leu Ile Ala	Ile Met Ala Thr Gly	Gly		
	500		505		510
Gly Thr Arg Ser	Met Thr Ser Met Tyr	Gly His Leu Leu Gly	Leu		
	515		520		525
Gln Lys Leu Asn	Leu Leu Asp Cys Ala	Ser Tyr Ile Thr Gly	Leu		
	530		535		540
Ser Gly Ala Thr	Trp Thr Met Ala Thr	Leu Tyr Arg Asp Pro	Asp		
	545		550		555
Trp Ser Ser Lys	Asn Leu Glu Pro Ala	Ile Phe Glu Ala Arg	Arg		
	560		565		570
His Val Val Lys	Asp Lys Leu Pro Ser	Leu Phe Pro Asp Gln	Leu		
	575		580		585
Arg Lys Phe Gln	Glu Glu Leu Arg Gln	Arg Ser Gln Glu Gly	Tyr		
	590		595		600
Arg Val Thr Phe	Thr Asp Phe Trp Gly	Leu Leu Ile Glu Thr	Cys		
	605		610		615
Leu Gly Asp Glu	Arg Asn Glu Cys Lys	Leu Ser Asp Gln Arg	Ala		
	620		625		630
Ala Leu Ser Cys	Gly Gln Asn Pro Leu	Pro Ile Tyr Leu Thr	Ile		
	635		640		645
Asn Val Lys Asp	Asp Val Ser Asn Gln	Asp Val Arg Trp Phe	Glu		
	650		655		660
Phe Ser Pro Tyr	Glu Val Gly Leu Gln	Lys Tyr Gly Ala Phe	Ile		
	665		670		675
Pro Ser Glu Leu	Phe Gly Ser Glu Phe	Phe Met Gly Arg Leu	Val		
	680		685		690
Lys Arg Ile Pro	Glu Ser Arg Ile Cys	Tyr Met Leu Gly Leu	Trp		
	695		700		705
Ser Ser Ile Phe	Ser Leu Asn Leu Leu	Asp Ala Trp Asn Leu	Ser		
	710		715		720
His Thr Ser Glu	Glu Phe Phe His Arg	Trp Thr Arg Glu Lys	Val		
	725		730		735
Gln Asp Ile Glu	Asp Glu Pro Ile Leu	Pro Glu Ile Pro Lys	Cys		
	740		745		750
Asp Ala Asn Ile	Leu Glu Thr Thr Val	Val Ile Pro Gly Ser	Trp		
	755		760		765
Leu Ser Asn Ser	Phe Arg Glu Ile Leu	Thr His Arg Ser Phe	Val		
	770		775		780
Ser Glu Phe His	Asn Phe Leu Ser Gly	Leu Gln Leu His Thr	Asn		
	785		790		795
Tyr Leu Gln Asn	Gly Gln Phe Ser Arg	Trp Lys Asp Thr Val	Leu		
	800		805		810
Asp Gly Phe Pro	Asn Gln Leu Thr Glu	Ser Ala Asn His Leu	Cys		
	815		820		825
Leu Leu Asp Thr	Ala Phe Phe Val Asn	Ser Ser Tyr Pro Pro	Leu		
	830		835		840
Leu Arg Pro Glu	Arg Lys Ala Asp Leu	Ile Ile His Leu Asn	Tyr		
	845		850		855
Cys Ala Gly Ser	Gln Thr Lys Pro Leu	Lys Gln Thr Cys Glu	Tyr		
	860		865		870
Cys Thr Val Gln	Asn Ile Pro Phe Pro	Lys Tyr Glu Leu Pro	Asp		
	875		880		885
Glu Asn Glu Asn	Leu Lys Glu Cys Tyr	Leu Met Glu Asn Pro	Gln		

	890		895		900
Glu Pro Asp Ala	Pro Ile Val Thr Phe	Phe Pro Leu Ile Asn Asp			
	905		910		915
Thr Phe Arg Lys	Tyr Lys Ala Pro Gly	Val Glu Arg Ser Pro Glu			
	920		925		930
Glu Leu Glu Gln	Gly Gln Val Asp Ile	Tyr Gly Pro Lys Thr Pro			
	935		940		945
Tyr Ala Thr Lys	Glu Leu Thr Tyr Thr	Glu Ala Thr Phe Asp Lys			
	950		955		960
Leu Val Lys Leu	Ser Glu Tyr Asn Ile	Leu Asn Asn Lys Asp Thr			
	965		970		975
Leu Leu Gln Ala	Leu Arg Leu Ala Val	Glu Lys Lys Lys Arg Leu			
	980		985		990
Lys Gly Gln Cys	Pro Ser				
	995				

&lt;210&gt; 2

&lt;211&gt; 372

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2884821CD1

&lt;400&gt; 2

Met Ser Leu Arg Cys	Gly Asp Ala Ala Arg Thr Leu Gly Pro Arg	
1	5	10 15
Val Phe Gly Arg Tyr	Phe Cys Ser Pro Val Arg Pro Leu Ser Ser	
	20	25 30
Leu Pro Asp Lys Lys	Lys Glu Leu Leu Gln Asn Gly Pro Asp Leu	
	35	40 45
Gln Asp Phe Val Ser	Gly Asp Leu Ala Asp Arg Ser Thr Trp Asp	
	50	55 60
Glu Tyr Lys Gly Asn	Leu Lys Arg Gln Lys Gly Glu Arg Leu Arg	
	65	70 75
Leu Pro Pro Trp Leu	Lys Thr Glu Ile Pro Met Gly Lys Asn Tyr	
	80	85 90
Asn Lys Leu Lys Asn	Thr Leu Arg Asn Leu Asn Leu His Thr Val	
	95	100 105
Cys Glu Glu Ala Arg	Cys Pro Asn Ile Gly Glu Cys Trp Gly Gly	
	110	115 120
Gly Glu Tyr Ala Thr	Ala Thr Ala Thr Ile Met Leu Met Gly Asp	
	125	130 135
Thr Cys Thr Arg Gly	Cys Arg Phe Cys Ser Val Lys Thr Ala Arg	
	140	145 150
Asn Pro Pro Pro Leu	Asp Ala Ser Glu Pro Tyr Asn Thr Ala Lys	
	155	160 165
Ala Ile Ala Glu Trp	Gly Leu Asp Tyr Val Val Leu Thr Ser Val	
	170	175 180
Asp Arg Asp Asp Met	Pro Asp Gly Gly Ala Glu His Ile Ala Lys	
	185	190 195
Thr Val Ser Tyr Leu	Lys Glu Arg Asn Pro Lys Ile Leu Val Glu	
	200	205 210
Cys Leu Thr Pro Asp	Phe Arg Gly Asp Leu Lys Ala Ile Glu Lys	
	215	220 225

Val	Ala	Leu	Ser	Gly	Leu	Asp	Val	Tyr	Ala	His	Asn	Val	Glu	Thr	230	235	240
Val	Pro	Glu	Leu	Gln	Ser	Lys	Val	Arg	Asp	Pro	Arg	Ala	Asn	Phe	245	250	255
Asp	Gln	Ser	Leu	Arg	Val	Leu	Lys	His	Ala	Lys	Lys	Val	Gln	Pro	260	265	270
Asp	Val	Ile	Ser	Lys	Thr	Ser	Ile	Met	Leu	Gly	Leu	Gly	Glu	Asn	275	280	285
Asp	Glu	Gln	Val	Tyr	Ala	Thr	Met	Lys	Ala	Leu	Arg	Glu	Ala	Asp	290	295	300
Val	Asp	Cys	Leu	Thr	Leu	Gly	Gln	Tyr	Met	Gln	Pro	Thr	Arg	Arg	305	310	315
His	Leu	Lys	Val	Glu	Glu	Tyr	Ile	Thr	Pro	Glu	Lys	Phe	Lys	Tyr	320	325	330
Trp	Glu	Lys	Val	Gly	Asn	Glu	Leu	Gly	Phe	His	Tyr	Thr	Ala	Ser	335	340	345
Gly	Pro	Leu	Val	Arg	Ser	Ser	Tyr	Lys	Ala	Gly	Glu	Phe	Phe	Leu	350	355	360
Lys	Asn	Leu	Val	Ala	Lys	Arg	Lys	Thr	Lys	Asp	Leu				365	370	

&lt;210&gt; 3

&lt;211&gt; 649

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 72852842CD1

&lt;400&gt; 3

Met	Glu	Met	Arg	Trp	Phe	Leu	Ser	Lys	Ile	Gln	Asp	Asp	Phe	Arg	1	5	10	15
Gly	Gly	Lys	Ile	Asn	Leu	Glu	Lys	Thr	Gln	Arg	Leu	Leu	Glu	Lys	20		25	30
Leu	Asp	Ile	Arg	Cys	Ser	Tyr	Ile	His	Val	Lys	Gln	Ile	Phe	Lys	35		40	45
Asp	Asn	Asp	Arg	Leu	Lys	Gln	Gly	Arg	Ile	Thr	Ile	Glu	Glu	Phe	50		55	60
Arg	Ala	Ile	Tyr	Arg	Ile	Ile	Thr	His	Arg	Glu	Glu	Ile	Ile	Glu	65		70	75
Ile	Phe	Asn	Thr	Tyr	Ser	Glu	Asn	Arg	Lys	Ile	Leu	Leu	Ala	Ser	80		85	90
Asn	Leu	Ala	Gln	Phe	Leu	Thr	Gln	Glu	Gln	Tyr	Ala	Ala	Glu	Met	95		100	105
Ser	Lys	Ala	Ile	Ala	Phe	Glu	Ile	Ile	Gln	Lys	Tyr	Glu	Pro	Ile	110		115	120
Glu	Glu	Val	Arg	Lys	Ala	His	Gln	Met	Ser	Leu	Glu	Gly	Phe	Thr	125		130	135
Arg	Tyr	Met	Asp	Ser	Arg	Glu	Cys	Leu	Leu	Phe	Lys	Asn	Glu	Cys	140		145	150
Arg	Lys	Val	Tyr	Gln	Asp	Met	Thr	His	Pro	Leu	Asn	Asp	Tyr	Phe	155		160	165
Ile	Ser	Ser	Ser	His	Asn	Thr	Tyr	Leu	Val	Ser	Asp	Gln	Leu	Leu	170		175	180
Gly	Pro	Ser	Asp	Leu	Trp	Gly	Tyr	Val	Ser	Ala	Leu	Val	Lys	Gly				



	185	190	195
Cys Arg Cys Leu	Glu Ile Asp Cys Trp	Asp Gly Ala Gln Asn	Glu
	200	205	210
Pro Val Val Tyr	His Gly Tyr Thr Leu	Thr Ser Lys Leu Leu	Phe
	215	220	225
Lys Thr Val Ile	Gln Ala Ile His Lys	Tyr Ala Phe Met Thr	Ser
	230	235	240
Asp Tyr Pro Val	Val Leu Ser Leu Glu	Asn His Cys Ser Thr	Ala
	245	250	255
Gln Gln Glu Val	Met Ala Asp Asn Leu	Gln Ala Thr Phe Gly	Glu
	260	265	270
Ser Leu Leu Ser	Asp Met Leu Asp Asp	Phe Pro Asp Thr Leu	Pro
	275	280	285
Ser Pro Glu Ala	Leu Lys Phe Lys Ile	Leu Val Lys Asn Lys	Lys
	290	295	300
Ile Gly Thr Leu	Lys Glu Thr His Glu	Arg Lys Gly Ser Asp	Lys
	305	310	315
Arg Gly Lys Val	Glu Glu Trp Glu Glu	Glu Val Ala Asp Gly	Glu
	320	325	330
Glu Glu Glu Glu	Glu Glu Glu Glu Glu	Glu Glu Glu Glu Glu	Asp
	335	340	345
Lys Phe Lys Glu	Ser Glu Val Leu Glu	Ser Val Leu Gly Asp	Asn
	350	355	360
Gln Asp Lys Glu	Thr Gly Val Lys Lys	Leu Pro Gly Val Met	Leu
	365	370	375
Phe Lys Lys Lys	Lys Thr Arg Lys Leu	Lys Ile Ala Leu Ala	Leu
	380	385	390
Ser Asp Leu Val	Ile Tyr Thr Lys Ala	Glu Lys Phe Lys Ser	Phe
	395	400	405
Gln His Ser Arg	Leu Tyr Gln Gln Phe	Asn Glu Asn Asn Ser	Ile
	410	415	420
Gly Glu Thr Gln	Ala Arg Lys Leu Ser	Lys Leu Arg Val His	Glu
	425	430	435
Phe Ile Phe His	Thr Arg Lys Phe Ile	Thr Arg Ile Tyr Pro	Lys
	440	445	450
Ala Thr Arg Ala	Asp Ser Ser Asn Phe	Asn Pro Gln Glu Phe	Trp
	455	460	465
Asn Ile Gly Cys	Gln Met Val Ala Leu	Asn Phe Gln Thr Pro	Gly
	470	475	480
Leu Pro Met Asp	Leu Gln Asn Gly Lys	Phe Leu Asp Asn Gly	Gly
	485	490	495
Ser Gly Tyr Ile	Leu Lys Pro His Phe	Leu Arg Glu Ser Lys	Ser
	500	505	510
Tyr Phe Asn Pro	Ser Asn Ile Lys Glu	Gly Met Pro Ile Thr	Leu
	515	520	525
Thr Ile Arg Leu	Ile Ser Gly Ile Gln	Leu Pro Leu Thr His	Ser
	530	535	540
Ser Ser Asn Lys	Gly Asp Ser Leu Val	Ile Ile Glu Val Phe	Gly
	545	550	555
Val Pro Asn Asp	Gln Met Lys Gln Gln	Thr Arg Val Ile Lys	Lys
	560	565	570
Asn Ala Phe Ser	Pro Arg Trp Asn Glu	Thr Phe Thr Phe Ile	Ile
	575	580	585
His Val Pro Glu	Leu Ala Leu Ile Arg	Phe Val Val Glu Gly	Gln
	590	595	600
Gly Leu Ile Ala	Gly Asn Glu Phe Leu	Gly Gln Tyr Thr Leu	Pro

	605		610		615
Leu Leu Cys Met Asn Lys Gly Tyr Arg	Arg	Ile Pro Leu Phe Ser			
	620		625		630
Arg Met Gly Glu Ser Leu Glu Pro Ala	Ser Leu Phe Val Tyr	Val			
	635		640		645
Trp Tyr Val Arg					

&lt;210&gt; 4

&lt;211&gt; 2020

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7484271CD1

&lt;400&gt; 4

Met Ser Gly Gly Leu Val Pro Ile Tyr Val	Ile Ala Gly Val Val
1	5 10 15
Thr Arg Lys Gly Arg Arg Gly Trp Asp	Ile Met Met Gln Leu Thr
	20 25 30
Leu Asn Thr Leu Phe Pro Val Val Ser Thr	Pro Ala Ile Thr Tyr
	35 40 45
Ile Val Thr Val Phe Thr Gly Asp Val Arg	Gly Ala Gly Thr Lys
	50 55 60
Ser Lys Ile Tyr Leu Val Met Tyr Gly Ala	Arg Gly Asn Lys Asn
	65 70 75
Ser Gly Lys Ile Phe Leu Glu Gly Gly Val	Phe Asp Arg Gly Arg
	80 85 90
Thr Asp Ile Phe His Ile Glu Leu Ala Val	Leu Leu Ser Pro Leu
	95 100 105
Ser Arg Val Ser Val Gly His Gly Asn Val	Gly Val Asn Arg Gly
	110 115 120
Trp Phe Cys Glu Lys Val Val Ile Leu Cys	Pro Phe Thr Gly Ile
	125 130 135
Gln Gln Thr Phe Pro Cys Ser Asn Trp Leu	Asp Glu Lys Lys Ala
	140 145 150
Asp Gly Leu Ile Glu Arg Gln Leu Tyr Glu	Met Val Ser Leu Arg
	155 160 165
Lys Lys Arg Leu Lys Lys Phe Pro Trp Ser	Leu Trp Val Trp Thr
	170 175 180
Thr Asp Leu Lys Lys Ala Gly Thr Asn Ser	Pro Ile Phe Ile Gln
	185 190 195
Ile Tyr Gly Gln Lys Gly Arg Thr Asp Glu	Ile Leu Leu Asn Pro
	200 205 210
Asn Asn Lys Trp Phe Lys Pro Gly Ile Ile	Glu Lys Phe Arg Ile
	215 220 225
Glu Leu Pro Asp Leu Gly Arg Phe Tyr Lys	Ile Arg Val Trp His
	230 235 240
Asp Lys Arg Ser Ser Gly Ser Gly Trp His	Leu Glu Arg Met Thr
	245 250 255
Leu Met Asn Thr Leu Asn Lys Asp Lys Tyr	Asn Phe Asn Cys Asn
	260 265 270
Arg Trp Leu Asp Ala Asn Glu Asp Asp Asn	Glu Ile Val Arg Glu
	275 280 285

Met Thr Ala Glu Gly Pro Thr Val Arg Arg Ile Met Gly Met Ala	290	295	300
Arg Tyr His Val Thr Val Cys Thr Gly Glu Leu Glu Gly Ala Gly	305	310	315
Thr Asp Ala Asn Val Tyr Leu Cys Leu Phe Gly Asp Val Gly Asp	320	325	330
Thr Gly Glu Arg Leu Leu Tyr Asn Cys Arg Asn Asn Thr Asp Leu	335	340	345
Phe Glu Lys Gly Asn Ala Asp Glu Phe Thr Ile Glu Ser Val Thr	350	355	360
Met Arg Asn Val Arg Arg Val Arg Ile Arg His Asp Gly Lys Gly	365	370	375
Ser Gly Ser Gly Trp Tyr Leu Asp Arg Val Leu Val Arg Glu Glu	380	385	390
Gly Gln Pro Glu Ser Asp Asn Val Glu Phe Pro Cys Leu Arg Trp	395	400	405
Leu Asp Lys Asp Lys Asp Asp Gly Gln Leu Val Arg Glu Leu Leu	410	415	420
Pro Ser Asp Ser Ser Ala Thr Leu Lys Asn Phe Arg Tyr His Ile	425	430	435
Ser Leu Lys Thr Gly Asp Val Ser Gly Ala Ser Thr Asp Ser Arg	440	445	450
Val Tyr Ile Lys Leu Tyr Gly Asp Lys Ser Asp Thr Ile Lys Gln	455	460	465
Val Leu Leu Val Ser Asp Asn Asn Leu Lys Asp Tyr Phe Glu Arg	470	475	480
Gly Arg Val Asp Glu Phe Thr Leu Glu Thr Leu Asn Ile Gly Asn	485	490	495
Ile Asn Arg Leu Val Ile Gly His Asp Ser Thr Gly Met His Ala	500	505	510
Ser Trp Phe Leu Gly Ser Val Gln Ile Arg Val Pro Arg Gln Gly	515	520	525
Lys Gln Tyr Thr Phe Pro Ala Asn Arg Trp Leu Asp Lys Asn Gln	530	535	540
Ala Asp Gly Arg Leu Glu Val Glu Leu Tyr Pro Ser Glu Val Val	545	550	555
Glu Ile Gln Lys Leu Val His Tyr Glu Val Glu Ile Trp Thr Gly	560	565	570
Asp Val Gly Gly Ala Gly Thr Ser Ala Arg Val Tyr Met Gln Ile	575	580	585
Tyr Gly Glu Lys Gly Lys Thr Glu Val Leu Phe Leu Ser Ser Arg	590	595	600
Ser Lys Val Phe Glu Arg Ala Ser Lys Asp Thr Phe Gln Leu Glu	605	610	615
Ala Ala Asp Val Gly Glu Val Tyr Lys Leu Arg Leu Gly His Thr	620	625	630
Gly Glu Gly Phe Gly Pro Ser Trp Phe Val Asp Thr Val Trp Leu	635	640	645
Arg His Leu Val Val Arg Glu Val Asp Leu Thr Pro Glu Glu Glu	650	655	660
Ala Arg Lys Lys Lys Glu Lys Asp Lys Leu Arg Gln Leu Leu Lys	665	670	675
Lys Glu Arg Leu Lys Ala Lys Leu Gln Arg Lys Lys Lys Lys Arg	680	685	690
Lys Gly Ser Asp Glu Glu Asp Glu Gly Glu Glu Glu Glu Ser Ser	695	700	705

Ser Ser Glu Glu Ser Ser Ser Glu Glu Glu Glu Met Glu Glu Glu	710	715	720
Glu Glu Glu Glu Glu Phe Gly Pro Gly Met Gln Glu Val Ile Glu	725	730	735
Gln His Lys Phe Glu Ala His Arg Trp Leu Ala Arg Gly Lys Glu	740	745	750
Asp Asn Glu Leu Val Val Glu Leu Val Pro Ala Gly Lys Pro Gly	755	760	765
Pro Glu Arg Asn Thr Tyr Glu Val Gln Val Val Thr Gly Asn Val	770	775	780
Pro Lys Ala Gly Thr Asp Ala Asn Val Tyr Leu Thr Ile Tyr Gly	785	790	795
Glu Glu Tyr Gly Asp Thr Gly Glu Arg Pro Leu Lys Lys Ser Asp	800	805	810
Lys Ser Asn Lys Phe Glu Gln Gly Gln Thr Asp Thr Phe Thr Ile	815	820	825
Tyr Ala Ile Asp Leu Gly Ala Leu Thr Lys Ile Arg Ile Arg His	830	835	840
Asp Asn Thr Gly Asn Arg Ala Gly Trp Phe Leu Asp Arg Ile Asp	845	850	855
Ile Thr Asp Met Asn Asn Glu Ile Thr Tyr Tyr Phe Pro Cys Gln	860	865	870
Arg Trp Leu Ala Val Glu Glu Asp Asp Gly Gln Leu Ser Arg Glu	875	880	885
Leu Leu Pro Val Asp Glu Ser Tyr Val Leu Pro Gln Ser Glu Glu	890	895	900
Gly Gly Gly Gly Gly Asp Asn Asn Pro Leu Asp Asn Leu Ala Leu	905	910	915
Glu Gln Lys Asp Lys Ser Thr Thr Phe Ser Val Thr Ile Lys Thr	920	925	930
Gly Val Lys Lys Asn Ala Gly Thr Asp Ala Asn Val Phe Ile Thr	935	940	945
Leu Phe Gly Thr Gln Asp Asp Thr Gly Met Thr Leu Leu Lys Ser	950	955	960
Ser Lys Thr Asn Ser Asp Lys Phe Glu Arg Asp Ser Ile Glu Ile	965	970	975
Phe Thr Val Glu Thr Leu Asp Leu Gly Asp Leu Trp Lys Val Arg	980	985	990
Leu Gly His Asp Asn Thr Gly Lys Ala Pro Gly Trp Phe Val Asp	995	1000	1005
Trp Val Glu Val Asp Ala Pro Ser Leu Gly Lys Cys Met Thr Phe	1010	1015	1020
Pro Cys Gly Arg Trp Leu Ala Lys Asn Glu Asp Asp Gly Ser Ile	1025	1030	1035
Ile Arg Asp Leu Phe His Ala Glu Leu Gln Thr Arg Leu Tyr Thr	1040	1045	1050
Pro Phe Val Pro Tyr Glu Ile Thr Leu Tyr Thr Ser Asp Val Phe	1055	1060	1065
Ala Ala Gly Thr Asp Ala Asn Ile Phe Ile Ile Ile Tyr Gly Cys	1070	1075	1080
Asp Ala Val Cys Thr Gln Gln Lys Tyr Leu Cys Thr Asn Lys Arg	1085	1090	1095
Glu Gln Lys Gln Phe Phe Glu Arg Lys Ser Ala Ser Arg Phe Ile	1100	1105	1110
Val Glu Leu Glu Asp Val Gly Glu Ile Ile Glu Lys Ile Arg Ile	1115	1120	1125

Gly His Asn Asn Thr Gly Met Asn Pro Gly Trp His Cys Ser His	1130	1135	1140
Val Asp Ile Arg Arg Leu Leu Pro Asp Lys Asp Gly Ala Glu Thr	1145	1150	1155
Leu Thr Phe Pro Cys Asp Arg Trp Leu Ala Thr Ser Glu Asp Asp	1160	1165	1170
Lys Lys Thr Ile Arg Glu Leu Val Pro Tyr Asp Ile Phe Thr Glu	1175	1180	1185
Lys Tyr Met Lys Asp Gly Ser Leu Arg Gln Val Tyr Lys Glu Val	1190	1195	1200
Glu Glu Pro Leu Asp Ile Val Leu Tyr Ser Val Gln Ile Phe Thr	1205	1210	1215
Gly Asn Ile Pro Gly Ala Gly Thr Asp Ala Lys Val Tyr Ile Thr	1220	1225	1230
Ile Tyr Gly Asp Leu Gly Asp Thr Gly Glu Arg Tyr Leu Gly Lys	1235	1240	1245
Ser Glu Asn Arg Thr Asn Lys Phe Glu Arg Gly Thr Ala Asp Thr	1250	1255	1260
Phe Ile Ile Glu Ala Ala Asp Leu Gly Val Ile Tyr Lys Ile Lys	1265	1270	1275
Leu Arg His Asp Asn Ser Lys Trp Cys Ala Asp Trp Tyr Val Glu	1280	1285	1290
Lys Val Glu Ile Trp Asn Asp Thr Asn Glu Asp Glu Phe Leu Phe	1295	1300	1305
Leu Cys Gly Arg Trp Leu Ser Leu Lys Lys Glu Asp Gly Arg Leu	1310	1315	1320
Glu Arg Leu Phe Tyr Glu Lys Glu Tyr Thr Gly Asp Arg Ser Ser	1325	1330	1335
Asn Cys Ser Ser Pro Ala Asp Phe Trp Glu Ile Ala Leu Ser Ser	1340	1345	1350
Lys Met Ala Asp Val Asp Ile Ser Thr Val Thr Gly Pro Met Ala	1355	1360	1365
Asp Tyr Val Gln Glu Gly Pro Ile Ile Pro Tyr Tyr Val Ser Val	1370	1375	1380
Thr Thr Gly Lys His Lys Asp Ala Ala Thr Asp Ser Arg Ala Phe	1385	1390	1395
Ile Phe Leu Ile Gly Glu Asp Asp Glu Arg Ser Lys Arg Ile Trp	1400	1405	1410
Leu Asp Tyr Pro Arg Gly Lys Arg Gly Phe Ser Arg Gly Ser Val	1415	1420	1425
Glu Glu Phe Tyr Val Ala Gly Leu Asp Val Gly Ile Ile Lys Lys	1430	1435	1440
Ile Glu Leu Gly His Asp Gly Ala Ser Pro Glu Ser Cys Trp Leu	1445	1450	1455
Val Glu Glu Leu Cys Leu Ala Val Pro Thr Gln Gly Thr Lys Tyr	1460	1465	1470
Met Leu Asn Cys Asn Cys Trp Leu Ala Lys Asp Arg Gly Asp Gly	1475	1480	1485
Ile Thr Ser Arg Val Phe Asp Leu Leu Asp Ala Met Val Val Asn	1490	1495	1500
Ile Gly Val Lys Val Leu Tyr Glu Met Thr Val Trp Thr Gly Asp	1505	1510	1515
Val Val Gly Gly Gly Thr Asp Ser Asn Ile Phe Met Thr Leu Tyr	1520	1525	1530
Gly Ile Asn Gly Ser Thr Glu Glu Met Gln Leu Asp Lys Lys Lys	1535	1540	1545

Ala Arg Phe Glu Arg Glu Gln Asn Asp Thr Phe Ile Met Glu Ile	1550	1555	1560
Leu Asp Ile Ala Pro Phe Thr Lys Met Arg Ile Arg Ile Asp Gly	1565	1570	1575
Leu Gly Ser Arg Pro Glu Trp Phe Leu Glu Arg Ile Leu Leu Lys	1580	1585	1590
Asn Met Asn Thr Gly Asp Leu Thr Met Phe Tyr Tyr Gly Asp Trp	1595	1600	1605
Leu Ser Gln Arg Lys Gly Lys Lys Thr Leu Val Cys Glu Met Cys	1610	1615	1620
Ala Val Ile Asp Glu Glu Glu Met Met Glu Trp Thr Ser Tyr Thr	1625	1630	1635
Val Ala Val Lys Thr Ser Asp Ile Leu Gly Ala Gly Thr Asp Ala	1640	1645	1650
Asn Val Phe Ile Ile Ile Phe Gly Glu Asn Gly Asp Ser Gly Thr	1655	1660	1665
Leu Ala Leu Lys Gln Ser Ala Asn Trp Asn Lys Phe Glu Arg Asn	1670	1675	1680
Asn Thr Asp Thr Phe Asn Phe Pro Asp Met Leu Ser Leu Gly His	1685	1690	1695
Leu Cys Lys Leu Arg Val Trp His Asp Asn Lys Gly Ile Phe Pro	1700	1705	1710
Gly Trp His Leu Ser Tyr Val Asp Val Lys Asp Asn Ser Arg Asp	1715	1720	1725
Glu Thr Phe His Phe Gln Cys Asp Cys Trp Leu Ser Lys Ser Glu	1730	1735	1740
Gly Asp Gly Gln Thr Val Arg Asp Phe Ala Cys Ala Asn Asn Lys	1745	1750	1755
Ile Cys Asp Glu Leu Glu Glu Thr Thr Tyr Glu Ile Val Ile Glu	1760	1765	1770
Thr Gly Asn Gly Gly Glu Thr Arg Glu Asn Val Trp Leu Ile Leu	1775	1780	1785
Glu Gly Arg Lys Asn Arg Ser Lys Glu Phe Leu Met Glu Asn Ser	1790	1795	1800
Ser Arg Gln Arg Ala Phe Arg Lys Gly Thr Thr Asp Thr Phe Glu	1805	1810	1815
Phe Asp Ser Ile Tyr Leu Gly Asp Ile Ala Ser Leu Cys Val Gly	1820	1825	1830
His Leu Ala Arg Glu Asp Arg Phe Ile Pro Lys Arg Glu Leu Ala	1835	1840	1845
Trp His Val Lys Thr Ile Thr Ile Thr Glu Met Glu Tyr Gly Asn	1850	1855	1860
Val Tyr Phe Phe Asn Cys Asp Cys Leu Ile Pro Leu Lys Arg Lys	1865	1870	1875
Arg Lys Tyr Phe Lys Val Phe Glu Val Thr Lys Thr Thr Glu Ser	1880	1885	1890
Phe Ala Ser Lys Val Gln Ser Leu Val Pro Val Lys Tyr Glu Val	1895	1900	1905
Ile Val Thr Thr Gly Tyr Glu Pro Gly Ala Gly Thr Asp Ala Asn	1910	1915	1920
Val Phe Val Thr Ile Phe Gly Ala Asn Gly Asp Thr Gly Lys Arg	1925	1930	1935
Glu Leu Lys Gln Lys Met Arg Asn Leu Phe Glu Arg Gly Ser Thr	1940	1945	1950
Asp Arg Phe Phe Leu Glu Thr Leu Glu Leu Gly Glu Leu Arg Lys	1955	1960	1965

Val Arg Leu Glu His Asp Ser Ser Gly Tyr Cys Ser Gly Trp Leu  
 1970 1975 1980  
 Val Glu Lys Val Glu Val Thr Asn Thr Ser Thr Gly Val Ala Thr  
 1985 1990 1995  
 Ile Phe Asn Cys Gly Arg Trp Leu Asp Lys Lys Arg Gly Asp Gly  
 2000 2005 2010  
 Leu Thr Trp Arg Asp Leu Phe Pro Ser Val  
 2015 2020

<210> 5  
 <211> 415  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7474074CD1

<400> 5  
 Met Met Tyr Thr Arg Asn Asn Leu Asn Cys Ala Glu Pro Leu Phe  
 1 5 10 15  
 Glu Gln Asn Asn Ser Leu Asn Val Asn Phe Asn Thr Gln Lys Lys  
 20 25 30  
 Thr Val Trp Leu Ile His Gly Tyr Arg Pro Val Gly Ser Ile Pro  
 35 40 45  
 Leu Trp Leu Gln Asn Phe Val Arg Ile Leu Leu Asn Glu Glu Asp  
 50 55 60  
 Met Asn Val Ile Val Val Asp Trp Ser Arg Gly Ala Thr Thr Phe  
 65 70 75  
 Ile Tyr Asn Arg Ala Val Lys Asn Thr Arg Lys Val Ala Val Ser  
 80 85 90  
 Leu Ser Val His Ile Lys Asn Leu Leu Lys His Gly Ala Ser Leu  
 95 100 105  
 Asp Asn Phe His Phe Ile Gly Val Ser Leu Gly Ala His Ile Ser  
 110 115 120  
 Gly Phe Val Gly Lys Ile Phe His Gly Gln Leu Gly Arg Ile Thr  
 125 130 135  
 Gly Leu Asp Pro Ala Gly Pro Arg Phe Ser Arg Lys Pro Pro Tyr  
 140 145 150  
 Ser Arg Leu Asp Tyr Thr Asp Ala Lys Phe Val Asp Val Ile His  
 155 160 165  
 Ser Asp Ser Asn Gly Leu Gly Ile Gln Glu Pro Leu Gly His Ile  
 170 175 180  
 Asp Phe Tyr Pro Asn Gly Gly Asn Lys Gln Pro Gly Cys Pro Lys  
 185 190 195  
 Ser Ile Phe Ser Gly Ile Gln Phe Ile Lys Cys Asn His Gln Arg  
 200 205 210  
 Ala Val His Leu Phe Met Ala Ser Leu Glu Thr Asn Cys Asn Phe  
 215 220 225  
 Ile Ser Phe Pro Cys Arg Ser Tyr Lys Asp Tyr Lys Thr Ser Leu  
 230 235 240  
 Cys Val Asp Cys Asp Cys Phe Lys Glu Lys Ser Cys Pro Arg Leu  
 245 250 255  
 Gly Tyr Gln Ala Lys Leu Phe Lys Gly Val Leu Lys Glu Arg Met  
 260 265 270  
 Glu Gly Arg Pro Leu Arg Thr Thr Val Phe Leu Asp Thr Ser Gly

275	280	285
Thr Tyr Pro Phe Cys Thr Tyr Tyr Phe	Val Leu Ser Ile Ile Val	
290	295	300
Pro Asp Lys Thr Met Met Asp Gly Ser	Phe Ser Phe Lys Leu Leu	
305	310	315
Asn Gln Leu Glu Met Ile Glu Glu Pro	Arg Leu Tyr Glu Lys Asn	
320	325	330
Lys Pro Phe Tyr Lys Leu Gln Glu Val	Lys Ile Leu Ala Gln Phe	
335	340	345
Tyr Asn Asp Phe Val Asn Ile Ser Ser	Ile Gly Leu Thr Tyr Phe	
350	355	360
Gln Ser Ser Asn Leu Gln Cys Ser Thr	Cys Thr Tyr Lys Ile Gln	
365	370	375
Ser Leu Met Leu Lys Ser Leu Thr Tyr	Pro Lys Arg Pro Pro Leu	
380	385	390
Cys Arg Tyr Asn Ile Val Leu Lys Glu	Arg Glu Glu Val Phe Leu	
395	400	405
Asn Pro Asn Thr Cys Thr Pro Lys Asn Thr		
410	415	

&lt;210&gt; 6

&lt;211&gt; 1152

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 72024970CD1

&lt;400&gt; 6

Met Ala Leu Pro Arg Gln Pro Asp Gln Gly Asn Gly Gly Leu Ala	
1 5 10 15	
Gly Gly Gly Thr Pro Leu Val Gly Gly Ser Val Val Leu Ser Ser	
20 25 30	
Glu Trp Gln Leu Gly Pro Leu Val Glu Arg Cys Met Gly Ala Met	
35 40 45	
Gln Glu Gly Met Gln Met Val Lys Leu Arg Gly Gly Ser Lys Gly	
50 55 60	
Leu Val Arg Phe Tyr Tyr Leu Asp Glu His Arg Ser Cys Ile Arg	
65 70 75	
Trp Arg Pro Ser Arg Lys Asn Glu Lys Ala Lys Ile Ser Ile Asp	
80 85 90	
Ser Ile Gln Glu Val Ser Glu Gly Arg Gln Ser Glu Val Phe Gln	
95 100 105	
Arg Tyr Pro Asp Gly Ser Phe Asp Pro Asn Cys Cys Phe Ser Ile	
110 115 120	
Tyr His Gly Ser His Arg Glu Ser Leu Asp Leu Val Ser Thr Ser	
125 130 135	
Ser Glu Val Ala Arg Thr Trp Val Thr Gly Leu Arg Tyr Leu Met	
140 145 150	
Ala Gly Ile Ser Asp Glu Asp Ser Leu Ala Arg Arg Gln Arg Thr	
155 160 165	
Arg Asp Gln Trp Leu Lys Gln Thr Phe Asp Glu Ala Asp Lys Asn	
170 175 180	
Gly Asp Gly Ser Leu Ser Ile Gly Glu Val Leu Gln Leu Leu His	
185 190 195	



Lys	Leu	Asn	Val	Asn	Leu	Pro	Arg	Gln	Arg	Val	Lys	Gln	Met	Phe	200	205	210
Arg	Glu	Ala	Asp	Thr	Asp	Asp	His	Gln	Gly	Thr	Leu	Gly	Phe	Glu	215	220	225
Glu	Phe	Cys	Ala	Phe	Tyr	Lys	Met	Met	Ser	Thr	Arg	Arg	Asp	Leu	230	235	240
Tyr	Leu	Leu	Met	Leu	Thr	Tyr	Ser	Asn	His	Lys	Asp	His	Leu	Asp	245	250	255
Ala	Ala	Ser	Leu	Gln	Arg	Phe	Leu	Gln	Val	Glu	Gln	Lys	Met	Ala	260	265	270
Gly	Val	Thr	Leu	Glu	Ser	Cys	Gln	Asp	Ile	Ile	Glu	Gln	Phe	Glu	275	280	285
Pro	Cys	Pro	Glu	Asn	Lys	Ser	Lys	Gly	Leu	Leu	Gly	Ile	Asp	Gly	290	295	300
Phe	Thr	Asn	Tyr	Thr	Arg	Ser	Pro	Ala	Gly	Asp	Ile	Phe	Asn	Pro	305	310	315
Glu	His	His	His	Val	His	Gln	Asp	Met	Thr	Gln	Pro	Leu	Ser	His	320	325	330
Tyr	Phe	Ile	Thr	Ser	Ser	His	Asn	Thr	Tyr	Leu	Val	Gly	Asp	Gln	335	340	345
Leu	Met	Ser	Gln	Ser	Arg	Val	Asp	Met	Tyr	Ala	Trp	Val	Leu	Gln	350	355	360
Ala	Gly	Cys	Arg	Cys	Val	Glu	Val	Asp	Cys	Trp	Asp	Gly	Pro	Asp	365	370	375
Gly	Glu	Pro	Ile	Val	His	His	Gly	Tyr	Thr	Leu	Thr	Ser	Lys	Ile	380	385	390
Leu	Phe	Lys	Asp	Val	Ile	Glu	Thr	Ile	Asn	Lys	Tyr	Ala	Phe	Ile	395	400	405
Lys	Asn	Glu	Tyr	Pro	Val	Ile	Leu	Ser	Ile	Glu	Asn	His	Cys	Ser	410	415	420
Val	Ile	Gln	Gln	Lys	Lys	Met	Ala	Gln	Tyr	Leu	Thr	Asp	Ile	Leu	425	430	435
Gly	Asp	Lys	Leu	Asp	Leu	Ser	Ser	Val	Ser	Ser	Glu	Asp	Ala	Thr	440	445	450
Thr	Leu	Pro	Ser	Pro	Gln	Met	Leu	Lys	Gly	Lys	Ile	Leu	Val	Lys	455	460	465
Gly	Lys	Lys	Leu	Pro	Ala	Asn	Ile	Ser	Glu	Asp	Ala	Glu	Glu	Gly	470	475	480
Glu	Val	Ser	Asp	Glu	Asp	Ser	Ala	Asp	Glu	Ile	Asp	Asp	Asp	Cys	485	490	495
Lys	Leu	Leu	Asn	Gly	Asp	Ala	Ser	Thr	Asn	Arg	Lys	Arg	Val	Glu	500	505	510
Asn	Thr	Ala	Lys	Arg	Lys	Leu	Asp	Ser	Leu	Ile	Lys	Glu	Ser	Lys	515	520	525
Ile	Arg	Asp	Cys	Glu	Asp	Pro	Asn	Asn	Phe	Ser	Val	Ser	Thr	Leu	530	535	540
Ser	Pro	Ser	Gly	Lys	Leu	Gly	Arg	Lys	Ser	Lys	Ala	Glu	Glu	Asp	545	550	555
Val	Glu	Ser	Gly	Glu	Asp	Ala	Gly	Ala	Ser	Arg	Arg	Asn	Gly	Arg	560	565	570
Leu	Val	Val	Gly	Ser	Phe	Ser	Arg	Arg	Lys	Lys	Lys	Gly	Ser	Lys	575	580	585
Leu	Lys	Lys	Ala	Ala	Ser	Val	Glu	Glu	Gly	Asp	Glu	Gly	Gln	Asp	590	595	600
Ser	Pro	Gly	Gly	Gln	Ser	Arg	Gly	Ala	Thr	Arg	Gln	Lys	Lys	Thr	605	610	615

Met Lys Leu Ser Arg Ala Leu Ser Asp	Leu Val Lys Tyr Thr Lys	
620	625	630
Ser Val Ala Thr His Asp Ile Glu Met	Glu Ala Ala Ser Ser Trp	
635	640	645
Gln Val Ser Ser Phe Ser Glu Thr Lys	Ala His Gln Ile Leu Gln	
650	655	660
Gln Lys Pro Ala Gln Tyr Leu Arg Phe	Asn Gln Gln Gln Leu Ser	
665	670	675
Arg Ile Tyr Pro Ser Ser Tyr Arg Val	Asp Ser Ser Asn Tyr Asn	
680	685	690
Pro Gln Pro Phe Trp Asn Ala Gly Cys	Gln Met Val Ala Leu Asn	
695	700	705
Tyr Gln Ser Glu Gly Arg Met Leu Gln	Leu Asn Arg Ala Lys Phe	
710	715	720
Ser Ala Asn Gly Gly Cys Gly Tyr Val	Leu Lys Pro Gly Cys Met	
725	730	735
Cys Gln Gly Val Phe Asn Pro Asn Ser	Glu Asp Pro Leu Pro Gly	
740	745	750
Gln Leu Lys Lys Gln Leu Val Leu Arg	Ile Ile Ser Gly Gln Gln	
755	760	765
Leu Pro Lys Pro Arg Asp Ser Met Leu	Gly Asp Arg Gly Glu Ile	
770	775	780
Ile Asp Pro Phe Val Glu Val Glu Ile	Ile Gly Leu Pro Val Asp	
785	790	795
Cys Ser Arg Glu Gln Thr Arg Val Val	Asp Asp Asn Gly Phe Asn	
800	805	810
Pro Thr Trp Glu Glu Thr Leu Val Phe	Met Val His Met Pro Glu	
815	820	825
Ile Ala Leu Val Arg Phe Leu Val Trp	Asp His Asp Pro Ile Gly	
830	835	840
Arg Asp Phe Ile Gly Gln Arg Thr Leu	Ala Phe Ser Ser Met Met	
845	850	855
Pro Gly Tyr Arg His Val Tyr Leu Glu	Gly Met Glu Glu Ala Ser	
860	865	870
Ile Phe Val His Val Ala Val Ser Asp	Ile Ser Gly Lys Val Lys	
875	880	885
Gln Ala Leu Gly Leu Lys Gly Leu Phe	Leu Arg Gly Pro Lys Pro	
890	895	900
Gly Ser Leu Asp Ser His Ala Ala Gly	Arg Pro Pro Ala Arg Pro	
905	910	915
Ser Val Ser Gln Arg Ile Leu Arg Arg	Thr Ala Ser Ala Pro Thr	
920	925	930
Lys Ser Gln Lys Pro Gly Arg Arg Gly	Phe Pro Glu Leu Val Leu	
935	940	945
Gly Thr Arg Asp Thr Gly Ser Lys Gly	Val Ala Asp Asp Val Val	
950	955	960
Pro Pro Gly Pro Gly Pro Ala Pro Glu	Ala Pro Ala Gln Glu Gly	
965	970	975
Pro Gly Ser Gly Ser Pro Arg Gly Lys	Ala Pro Ala Ala Val Ala	
980	985	990
Glu Lys Ser Pro Val Arg Val Arg Pro	Pro Arg Val Leu Asp Gly	
995	1000	1005
Pro Gly Pro Ala Gly Met Ala Ala Thr	Cys Met Lys Cys Val Val	
1010	1015	1020
Gly Ser Cys Ala Gly Val Asn Thr Gly	Gly Pro Gln Arg Glu Arg	
1025	1030	1035

```

Pro Pro Ser Pro Gly Pro Ala Ser Arg Gln Ala Ala Ile Arg Gln
      1040                      1045                      1050
Gln Pro Arg Ala Arg Ala Asp Ser Leu Gly Ala Pro Cys Cys Gly
      1055                      1060                      1065
Leu Asp Pro His Ala Ile Pro Gly Arg Ser Arg Glu Ala Pro Lys
      1070                      1075                      1080
Gly Pro Gly Ala Trp Arg Gln Gly Pro Gly Gly Ser Gly Ser Met
      1085                      1090                      1095
Ser Ser Asp Ser Ser Ser Pro Asp Ser Pro Gly Ile Pro Glu Arg
      1100                      1105                      1110
Ser Pro Arg Trp Pro Glu Gly Ala Cys Arg Gln Pro Gly Ala Leu
      1115                      1120                      1125
Gln Gly Glu Met Ser Ala Leu Phe Ala Gln Lys Leu Glu Glu Ile
      1130                      1135                      1140
Arg Ser Lys Ser Pro Met Phe Ser Ala Val Arg Asn
      1145                      1150

```

&lt;210&gt; 7

&lt;211&gt; 1294

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6131380CD1

&lt;400&gt; 7

```

Met Ile Ile Lys Glu Tyr Arg Ile Pro Leu Pro Met Thr Val Glu
  1              5              10              15
Glu Tyr Arg Ile Ala Gln Leu Tyr Met Ile Gln Lys Lys Ser Arg
      20              25              30
Asn Glu Thr Tyr Gly Glu Gly Ser Gly Val Glu Ile Leu Glu Asn
      35              40              45
Arg Pro Tyr Thr Asp Gly Pro Gly Gly Ser Gly Gln Tyr Thr His
      50              55              60
Lys Val Tyr His Val Gly Met His Ile Pro Ser Trp Phe Arg Ser
      65              70              75
Ile Leu Pro Lys Ala Ala Leu Arg Val Val Glu Glu Ser Trp Asn
      80              85              90
Ala Tyr Pro Tyr Thr Arg Thr Arg Phe Thr Cys Pro Phe Val Glu
      95              100             105
Lys Phe Ser Ile Asp Ile Glu Thr Phe Tyr Lys Thr Asp Ala Gly
      110             115             120
Glu Asn Pro Asp Val Phe Asn Leu Ser Pro Val Glu Lys Asn Gln
      125             130             135
Leu Thr Ile Asp Phe Ile Asp Ile Val Lys Asp Pro Val Pro His
      140             145             150
Asn Glu Tyr Lys Thr Glu Glu Asp Pro Lys Leu Phe Gln Ser Thr
      155             160             165
Lys Thr Gln Arg Gly Pro Leu Ser Glu Asn Trp Ile Glu Glu Tyr
      170             175             180
Lys Lys Gln Val Phe Pro Ile Met Cys Ala Tyr Lys Leu Cys Lys
      185             190             195
Val Glu Phe Arg Tyr Trp Gly Met Gln Ser Lys Ile Glu Arg Phe
      200             205             210
Ile His Asp Thr Gly Leu Arg Arg Val Met Val Arg Ala His Arg

```

	215		220		225
Gln Ala Trp Cys	Trp Gln Asp Glu Trp	Tyr Gly Leu Ser Met	Glu		
	230		235		240
Asn Ile Arg Glu	Leu Glu Lys Glu Ala	Gln Leu Met Leu Ser	Arg		
	245		250		255
Lys Met Ala Gln	Phe Asn Glu Asp Gly	Glu Glu Ala Thr Glu	Leu		
	260		265		270
Val Lys His Glu	Ala Val Ser Asp Gln	Thr Ser Gly Glu Pro	Pro		
	275		280		285
Glu Pro Ser Ser	Ser Asn Gly Glu Pro	Leu Val Gly Arg Gly	Leu		
	290		295		300
Lys Lys Gln Trp	Ser Thr Ser Ser Lys	Ser Ser Arg Ser Ser	Lys		
	305		310		315
Arg Gly Ala Ser	Pro Ser Arg His Ser	Ile Ser Glu Trp Arg	Met		
	320		325		330
Gln Ser Ile Ala	Arg Asp Ser Asp Glu	Ser Ser Asp Asp Glu	Phe		
	335		340		345
Phe Asp Ala His	Glu Asp Leu Ser Asp	Thr Glu Glu Met Phe	Pro		
	350		355		360
Lys Asp Ile Thr	Lys Trp Ser Ser Asn	Asp Leu Met Asp Lys	Ile		
	365		370		375
Glu Ser Pro Glu	Pro Glu Asp Thr Gln	Asp Gly Leu Tyr Arg	Gln		
	380		385		390
Gly Ala Pro Glu	Phe Arg Val Ala Ser	Ser Val Glu Gln Leu	Asn		
	395		400		405
Ile Ile Glu Asp	Glu Val Ser Gln Pro	Leu Ala Ala Pro Pro	Ser		
	410		415		420
Lys Ile His Val	Leu Leu Val Leu	His Gly Gly Thr Ile	Leu		
	425		430		435
Asp Thr Gly Ala	Gly Asp Pro Ser Ser	Lys Lys Gly Asp Ala	Asn		
	440		445		450
Thr Ile Ala Asn	Val Phe Asp Thr Val	Met Arg Val His Tyr	Pro		
	455		460		465
Ser Ala Leu Gly	Arg Leu Ala Ile Arg	Leu Val Pro Cys Pro	Pro		
	470		475		480
Val Cys Ser Asp	Ala Phe Ala Leu Val	Ser Asn Leu Ser Pro	Tyr		
	485		490		495
Ser His Asp Glu	Gly Cys Leu Ser Ser	Ser Gln Asp His Ile	Pro		
	500		505		510
Leu Ala Ala Leu	Pro Leu Leu Ala Thr	Ser Ser Pro Gln Tyr	Gln		
	515		520		525
Glu Ala Val Ala	Thr Val Ile Gln Arg	Ala Asn Leu Ala Tyr	Gly		
	530		535		540
Asp Phe Ile Lys	Ser Gln Glu Gly Met	Thr Phe Asn Gly Gln	Val		
	545		550		555
Cys Leu Ile Gly	Asp Cys Val Gly Gly	Ile Leu Ala Phe Asp	Ala		
	560		565		570
Leu Cys Tyr Ser	Asn Gln Pro Val Ser	Glu Ser Gln Ser Ser	Ser		
	575		580		585
Arg Arg Gly Ser	Val Val Ser Met Gln	Asp Asn Asp Leu Leu	Ser		
	590		595		600
Pro Gly Ile Leu	Met Asn Ala Ala His	Cys Cys Gly Gly Gly	Gly		
	605		610		615
Gly Gly Gly Gly	Gly Gly Gly Ser Ser	Gly Gly Gly Gly Ser	Ser		
	620		625		630
Gly Gly Ser Ser	Leu Glu Ser Ser Arg	His Leu Ser Arg Ser	Asn		

Val Asp Ile Pro	Arg Ser Asn Gly Thr	Glu Asp Pro Lys Arg Gln	635	640	645
Leu Pro Arg Lys	Arg Ser Asp Ser Ser Thr	Tyr Glu Leu Asp Thr	650	655	660
Ile Gln Gln His	Gln Ala Phe Leu Ser Ser	Leu His Ala Ser Val	665	670	675
Leu Arg Thr Glu	Pro Cys Ser Arg His Ser	Ser Ser Ser Thr Met	680	685	690
Leu Asp Gly Thr	Gly Ala Leu Gly Arg Phe	Asp Phe Glu Ile Thr	695	700	705
Asp Leu Phe Leu	Phe Gly Cys Pro Leu Gly	Leu Val Leu Ala Leu	710	715	720
Arg Lys Thr Val	Ile Pro Ala Leu Asp Val	Phe Gln Leu Arg Pro	725	730	735
Ala Cys Gln Gln	Val Tyr Asn Leu Phe His	Pro Ala Asp Pro Ser	740	745	750
Ala Ser Arg Leu	Glu Pro Leu Leu Glu Arg	Arg Phe His Ala Leu	755	760	765
Pro Pro Phe Ser	Val Pro Arg Tyr Gln Arg	Tyr Pro Leu Gly Asp	770	775	780
Gly Cys Ser Thr	Leu Leu Asp Val Leu Gln	Thr His Asn Ala Ala	785	790	795
Phe Gln Glu His	Gly Ala Pro Ser Ser Pro	Gly Thr Ala Pro Ala	800	805	810
Ser Arg Gly Phe	Arg Arg Ala Ser Glu Ile	Ser Ile Ala Ser Gln	815	820	825
Val Ser Gly Met	Ala Glu Ser Tyr Thr Ala	Ser Ser Ile Ala Gln	830	835	840
Val Ala Ala Lys	Trp Trp Gly Gln Lys Arg	Ile Asp Tyr Ala Leu	845	850	855
Tyr Cys Pro Asp	Ala Leu Thr Ala Phe Pro	Thr Val Ala Leu Pro	860	865	870
His Leu Phe His	Ala Ser Tyr Trp Glu Ser	Thr Asp Val Val Ser	875	880	885
Phe Leu Leu Arg	Gln Val Met Arg His Asp	Asn Ser Ser Ile Leu	890	895	900
Glu Leu Asp Gly	Lys Glu Val Ser Val Phe	Thr Pro Ser Lys Pro	905	910	915
Arg Glu Lys Trp	Gln Arg Lys Arg Thr His	Val Lys Leu Arg Asn	920	925	930
Val Thr Ala Asn	His Arg Ile Asn Asp Ala	Leu Ala Asn Glu Asp	935	940	945
Gly Pro Gln Val	Leu Thr Gly Arg Phe Met	Tyr Gly Pro Leu Asp	950	955	960
Met Val Thr Leu	Thr Gly Glu Lys Val Asp	Val His Ile Met Thr	965	970	975
Gln Pro Pro Ser	Gly Glu Trp Leu Tyr Leu	Asp Thr Leu Val Thr	980	985	990
Asn Asn Ser Gly	Arg Val Ser Tyr Thr Ile	Pro Glu Ser His Arg	995	1000	1005
Leu Gly Val Gly	Val Tyr Pro Ile Lys Met	Val Val Arg Gly Asp	1010	1015	1020
His Thr Phe Ala	Asp Ser Tyr Ile Thr Val	Leu Pro Lys Gly Thr	1025	1030	1035
Glu Phe Val Val	Phe Ser Ile Asp Gly Ser	Phe Ala Ala Ser Val	1040	1045	1050

1055	1060	1065
Ser Ile Met Gly Ser Asp Pro Lys Val Arg Ala Gly Ala Val Asp		
1070	1075	1080
Val Val Arg His Trp Gln Asp Leu Gly Tyr Leu Ile Ile Tyr Val		
1085	1090	1095
Thr Gly Arg Pro Asp Met Gln Lys Gln Arg Val Val Ala Trp Leu		
1100	1105	1110
Ala Gln His Asn Phe Pro His Gly Val Val Ser Phe Cys Asp Gly		
1115	1120	1125
Leu Val His Asp Pro Leu Arg His Lys Ala Asn Phe Leu Lys Leu		
1130	1135	1140
Leu Ile Ser Glu Leu His Leu Arg Val His Ala Ala Tyr Gly Ser		
1145	1150	1155
Thr Lys Asp Val Ala Val Tyr Ser Ala Ile Ser Leu Ser Pro Met		
1160	1165	1170
Gln Ile Tyr Ile Val Gly Arg Pro Thr Lys Lys Leu Gln Gln Gln		
1175	1180	1185
Cys Gln Phe Ile Thr Asp Gly Tyr Ala Ala His Leu Ala Gln Leu		
1190	1195	1200
Lys Tyr Ser His Arg Ala Arg Pro Ala Arg Asn Thr Ala Thr Arg		
1205	1210	1215
Met Ala Leu Arg Lys Gly Ser Phe Gly Leu Pro Gly Gln Gly Asp		
1220	1225	1230
Phe Leu Arg Ser Arg Asn His Leu Leu Arg Thr Ile Ser Ala Gln		
1235	1240	1245
Pro Ser Gly Pro Ser His Arg His Glu Arg Thr Gln Ser Gln Ala		
1250	1255	1260
Asp Gly Glu Gln Arg Gly Gln Arg Ser Met Ser Val Ala Ala Gly		
1265	1270	1275
Cys Trp Gly Arg Ala Met Thr Gly Arg Leu Glu Pro Gly Ala Ala		
1280	1285	1290
Ala Gly Pro Lys		

&lt;210&gt; 8

&lt;211&gt; 77

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 643681CD1

&lt;400&gt; 8

Met Asp Met Val Arg Trp Cys Gly Glu Asp Val Arg Lys Leu Glu		
1	5	10
Val Phe Ile Thr Ser Gln Gly Ala Ser Glu Tyr Arg Gly Lys Lys		
20	25	30
Thr Thr Lys Arg Gln Ala Gln Gly Glu Ser Thr Ile Lys Asp Ile		
35	40	45
Pro Met Pro Ala Ser Ile Ala Ala Pro Ala Leu Leu Ala Gly His		
50	55	60
Leu Pro Gln Leu His Leu Pro Ser Lys Leu Phe Asn Phe His Thr		
65	70	75
Val Ser		

<210> 9  
 <211> 576  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 6897474CD1

<400> 9

Met	Ala	Gln	Gly	Val	Leu	Trp	Ile	Leu	Leu	Gly	Leu	Leu	Leu	Trp	1	5	10	15
Ser	Asp	Pro	Gly	Thr	Ala	Ser	Leu	Pro	Leu	Leu	Met	Asp	Ser	Val	20	25	30	
Ile	Gln	Ala	Leu	Ala	Glu	Leu	Glu	Gln	Lys	Val	Pro	Ala	Ala	Lys	35	40	45	
Thr	Arg	His	Thr	Ala	Ser	Ala	Trp	Leu	Met	Ser	Ala	Pro	Asn	Ser	50	55	60	
Gly	Pro	His	Asn	Arg	Leu	Tyr	His	Phe	Leu	Leu	Gly	Ala	Trp	Ser	65	70	75	
Leu	Asn	Ala	Thr	Glu	Leu	Asp	Pro	Cys	Pro	Leu	Ser	Pro	Glu	Leu	80	85	90	
Leu	Gly	Leu	Thr	Lys	Glu	Val	Ala	Arg	His	Asp	Val	Arg	Glu	Gly	95	100	105	
Lys	Glu	Tyr	Gly	Val	Val	Leu	Ala	Pro	Asp	Gly	Ser	Thr	Val	Ala	110	115	120	
Val	Glu	Pro	Leu	Leu	Ala	Gly	Leu	Glu	Ala	Gly	Leu	Gln	Gly	Arg	125	130	135	
Arg	Val	Ile	Asn	Leu	Pro	Leu	Asp	Ser	Met	Ala	Ala	Pro	Trp	Glu	140	145	150	
Thr	Gly	Asp	Thr	Phe	Pro	Asp	Val	Val	Ala	Ile	Ala	Pro	Asp	Val	155	160	165	
Arg	Ala	Thr	Ser	Ser	Pro	Gly	Leu	Arg	Asp	Gly	Ser	Pro	Asp	Val	170	175	180	
Thr	Thr	Ala	Asp	Ile	Gly	Ala	Asn	Thr	Pro	Asp	Ala	Thr	Lys	Gly	185	190	195	
Cys	Pro	Asp	Val	Gln	Ala	Ser	Leu	Pro	Asp	Ala	Lys	Ala	Lys	Ser	200	205	210	
Pro	Pro	Thr	Met	Val	Asp	Ser	Leu	Leu	Ala	Val	Thr	Leu	Ala	Gly	215	220	225	
Asn	Leu	Gly	Leu	Thr	Phe	Leu	Arg	Gly	Ser	Gln	Thr	Gln	Ser	His	230	235	240	
Pro	Asp	Leu	Gly	Thr	Glu	Gly	Cys	Trp	Asp	Gln	Leu	Ser	Ala	Pro	245	250	255	
Arg	Thr	Phe	Thr	Leu	Leu	Asp	Pro	Lys	Ala	Ser	Leu	Leu	Thr	Met	260	265	270	
Ala	Phe	Leu	Asn	Gly	Ala	Leu	Asp	Gly	Val	Ile	Leu	Gly	Asp	Tyr	275	280	285	
Leu	Ser	Arg	Thr	Pro	Glu	Pro	Arg	Pro	Ser	Leu	Ser	His	Leu	Leu	290	295	300	
Ser	Gln	Tyr	Tyr	Gly	Ala	Gly	Val	Ala	Arg	Asp	Pro	Gly	Phe	Arg	305	310	315	
Ser	Asn	Phe	Arg	Arg	Gln	Asn	Gly	Ala	Ala	Leu	Thr	Ser	Ala	Ser	320	325	330	
Ile	Leu	Ala	Gln	Gln	Val	Trp	Gly	Thr	Leu	Val	Leu	Leu	Gln	Arg	335	340	345	

Leu	Glu	Pro	Val	His	Leu	Gln	Leu	Gln	Cys	Met	Ser	Gln	Glu	Gln
				350					355					360
Leu	Ala	Gln	Val	Ala	Ala	Asn	Ala	Thr	Lys	Glu	Phe	Thr	Glu	Ala
				365					370					375
Phe	Leu	Gly	Cys	Pro	Ala	Ile	His	Pro	Arg	Cys	Arg	Trp	Gly	Ala
				380					385					390
Ala	Pro	Tyr	Arg	Gly	Arg	Pro	Lys	Leu	Leu	Gln	Leu	Pro	Leu	Gly
				395					400					405
Phe	Leu	Tyr	Val	His	His	Thr	Tyr	Val	Pro	Ala	Pro	Pro	Cys	Thr
				410					415					420
Asp	Phe	Thr	Arg	Cys	Ala	Ala	Asn	Met	Arg	Ser	Met	Gln	Arg	Tyr
				425					430					435
His	Gln	Asp	Thr	Gln	Gly	Trp	Gly	Asp	Ile	Gly	Tyr	Ser	Phe	Val
				440					445					450
Val	Gly	Ser	Asp	Gly	Tyr	Val	Tyr	Glu	Gly	Arg	Gly	Trp	His	Trp
				455					460					465
Val	Gly	Ala	His	Thr	Leu	Gly	His	Asn	Ser	Arg	Gly	Phe	Gly	Val
				470					475					480
Ala	Ile	Val	Gly	Asn	Tyr	Thr	Ala	Ala	Leu	Pro	Thr	Glu	Ala	Ala
				485					490					495
Leu	Arg	Thr	Val	Arg	Asp	Thr	Leu	Pro	Ser	Cys	Ala	Val	Arg	Ala
				500					505					510
Gly	Leu	Leu	Arg	Pro	Asp	Tyr	Ala	Leu	Leu	Gly	His	Arg	Gln	Leu
				515					520					525
Val	Arg	Thr	Asp	Cys	Pro	Gly	Asp	Ala	Leu	Phe	Asp	Leu	Leu	Arg
				530					535					540
Thr	Trp	Pro	His	Phe	Thr	Ala	Thr	Val	Lys	Pro	Arg	Pro	Ala	Arg
				545					550					555
Ser	Val	Ser	Lys	Arg	Ser	Arg	Arg	Glu	Pro	Pro	Pro	Arg	Thr	Leu
				560					565					570
Pro	Ala	Thr	Asp	Leu	Gln									
				575										

&lt;210&gt; 10

&lt;211&gt; 3879

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7472774CB1

&lt;400&gt; 10

```

aggtccctgg ccacagctcc tgggggtacca agccatgaaa ctgaagtgga gttggggagcg 60
acggtcgcat cctcctagag gggcatctat gagccatgac ctctataagc tgaagagata 120
gagctttccc aaattatggc gggctagtcc tacagtcagtg tgggtccagt gtcctcttct 180
tgccaccacac tgtgcccttg aaggcctggt cattctgagt ggctgggggc tacagactgc 240
tgaccccaaaa gaccagagcc ctgcggggtcc ctgtatttct atgacctgaa gacctgtgat 300
ttcttttgata tgaagagatc taggcccacg caccctatct gtctaccac tcaaaccact 360
cccagagcaa tcccagctac tgccaagctg tggccaggaa ggtggagctc tgagtcagag 420
tataagtccc tgatcttgcc acccagctgg agagctgccg tgatgctcct gaggcagatg 480
cacgccaggg tctccactc cctgccagac ccatgccaa gagaagacag caggccctcg 540
gccacctgtg ccttgaaggc tccccagact tcatgggatg gtttgctgag ggaggggctg 600
tctccatgcc acctgttgac agtgagggtc atccggatga aaaatgtccg gcaggtgat 660
atgcaaccag taggtataga gctggcacc tgcctgcagg ctcccagcgt accggagaca 720
gacctgaagg gtgtggtcca ggcccggggt gggggggcca gtgttctgga aaagccaagg 780

```



```

gaaggggttca agaggggtga gcaggttcct gtgagccaga cagactgttt tgtgagcctc 840
tggctgcccc ccgcctctca gaagaagctg aggacaagga ccatctccaa ctgccccaaat 900
ccagagtggga atgaaagctt caacttccag atccagagcc gagtgaagaa cgtgctagag 960
ttgagtgtct gtgatgaaga cacagtga caagatgacc atctcctgac agttctctat 1020
gacctcacca agctctgttt ccgaaagaaa acccacgtga agtttccact caaccgcag 1080
ggcatggaag agctggaggt ggagttcctg ctggaggaga gtccctctcc acctgagacc 1140
ctcgtcacca atggcgtgct ggtggaatt atcttctgg gttcctgtag ctccagaggt 1200
cacggctggc tgcgtctctc aggggaacag gaccaaggga gaaaacagt ggcccagctt 1260
ggtctctgtc ctatctgac ctctgcagga gttagactaa acgaggccag ccaaattggg 1320
cacaggcagc actggggcac gagctggggc ttctgtacag agggaggagt gaaggacctc 1380
ctggtgatgg tgaacgaatc ctttgagaac acccagcgtg tccggccctg cttggaacct 1440
tgctgcccc cctctgcctg cttccaaacc gctgcctgct tccactacce caagtacttc 1500
cagtcaccag tgcacgtgga agtgcccaag agtactgga gctgtgggt ttgctgccgc 1560
tctcgcaaga agggcccat cagccagccc ctgcactgcc ttccgatgg tcaggtgatg 1620
accctgcctg tgggtgagag ttatgaatta cacatgaagt ctacaccctg ccctgagaca 1680
ctggacgtgc ggctgggctt cagcctgtgc ccagcagagc tggagtctt gcagaagcgg 1740
aaggtcgtgg tggccaaggc cctgaagcag gtgctgcagc tggaggaaga cctgcaggag 1800
gacgaggtgc cgtgatagc catcatggcc actgggggtg gaacaagatc catgacctcc 1860
atgtatggcc acctgctggg gctgcagaag ctgaacctcc tggactgtgc cagctacatc 1920
actggtctat caggggccac ctggaccatg gctacctgt accgtgacct tgactggtcc 1980
tccaaaaact tggagcctgc tatctttgag gctcggagac atgtggtaaa ggacaagcta 2040
ccctccctgt tcccagacca gctccgcaaa ttccaggagg agctccggca gcgcagccag 2100
gaaggctaca gggtcacctt tacagacttc tggggcctgc tgatagagac ctgcctgggg 2160
gacgagagaa atgaatgcaa actgtcagat cagcgtgctg ctttgagctg cggccagAAC 2220
ccctgcccc tctacctac catcaatgct aaggatgatg taagcaacca ggatgtcaga 2280
tggttcgagt tctccccta cgaggtgggc ctgcagaagt atggggcctt catccctcc 2340
gagctcttcg gctccgagtt cttcatgggg cggctgggtg agaggatccc ggagtctcga 2400
atctgtctaca tgctaggcct gtggagcagc atcttctccc tgaacctgct ggatgctgg 2460
aacctgtcac acacctcgga ggagttttcc cacagggtgga caaggagaa agtgcaggac 2520
atcgaagacg agccgatcct gcctgaaatc cccaaatgtg atgctaacat cctggagacc 2580
acggtagtga tcccagggtc atggctgtcc aattcttccc gagaaatcct taccatcgg 2640
tccttcgtgt ctgagtttca caacttctg tctgggctgc agctgcacac caactacctc 2700
cagaatggcc agttctctag gtggaagac acagtgttag atggtttccc aaaccagctg 2760
accgagtcgg cgaaccacct gtgcctgctg gacactgcgt tctttgtcaa ctccagctac 2820
ccgccccctc tcaggccaga gcgaaaagcc gatctcatca tccacctcaa ctactgtgt 2880
gggtcccaga caaagccct gaaacaaacc tgtgagtact gactgtgca gaacatcccc 2940
ttccccaaat acgagctgcc agatgagaat gaaaatctca agaatgcta cctgatggag 3000
aaccgccagg aaccgatgc ccccatcgtg actttcttcc cactcatcaa tgacacttcc 3060
cgaaaataca aggcaccagg ttagagcgga agccctgagg agctggagca gggccagggt 3120
gacatttatg gtcccaaac tccctatgcc accaaggagc tgacatacac agaggccacc 3180
tttgacaagc tgggtgaaact ctcagagtat aacatcctga ataataagga cactctctcc 3240
caggctctgc ggctcgcagt ggagaagaag aagcgcctga agggccagtg tccctcctag 3300
gccccaggga gcctccccctg ttctgtgtca gcttctacca tcagaggtgc aggaccctc 3360
agggctgacc aggttactac gcagccagct ctgctctccg gcaatgggtg tgagcagggt 3420
ggcctgggct ttctaacgaa aagtaaaaaa ttttaaaaag ttgagaaagt cagaaagaga 3480
gagagaggag ctctgttggg gttttatacc cactagagtt tcttcaagtg ctccctata 3540
gagaagggtg tctcatagcc acaggctccc acacatctgt ggagaggaaa agcctgggga 3600
agaggctggg cccccagaaa cctcgactca gaggcagagc ccagggtgg cagccctcct 3660
ctctctgtcc tctacctgt gtggcggggc tagggaaatg cacagaagga cctgagaggc 3720
actcggcgtt tctactgaaa aacacttcaa aatttaaggc aattctagtc ttgtgatttt 3780
tggttttttt tagacggagt ctcactctgt tgcccaggct ggagtgcaat ggcgcgatct 3840
cggctcactg caacctctgc ctcccaggtt caagcaatt 3879

```

&lt;210&gt; 11

&lt;211&gt; 1623

&lt;212&gt; DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2884821CB1

<400> 11

```

gcttggctgc ttgtcataaa tggagcgacg taatttcgac ctgtcctttc ccgggagtta 60
gcgatccctc aaccctgca ctgcgctagt cctaaagagg aaatgtctct acgctgcggg 120
gatgcagccc gcaccctggg gccccgggta tttgggagat atttttgcag cccagtcaga 180
ccgttaagct ccttgccaga taaaaaaaag gaactcctac agaattggacc agaccttcaa 240
gattttgtat ctggtgatct tgacagacagg agcacctggg atgaatataa aggaaaccta 300
aaacgccaga aaggagaaaag gttaagacta cctccatggc taaagacaga gattcccatg 360
gggaaaaaatt acaataaaact gaaaaaatac ttgcgggaatt taaatctcca tacagtatgt 420
gaggaagctc gatgtcccaa tattggagag tgttggggag gtggagaata tgccaccgcc 480
acagccacga tcatgttgat ggggtgacaca tgtacaagag gttgcagatt ttgttctgtt 540
aagactgcaa gaaatcctcc tccactggat gccagtggc cctacaatac tgcaaggcca 600
attgcagaat ggggtctgga ttatgttgtc ctgacatctg tggatcgaga tgatatgcct 660
gatggggggag ctgaacacat tgcaagacc gtatcatatt taaaggaaag gaatccaaaa 720
atccttgttg agtgtcttac tcttgatttt cgaggtgatc tcaaagcaat agaaaaagtt 780
gctctgtcag gattagatgt gtatgcacat aatgtagaaa cagtcccggg attacagagt 840
aaggttcgtg atcctcgggc caattttgat cagtcctac gtgtactgaa acatgccaag 900
aaggttcagc ctgatgttat ttctaaaaca tctataatgt tgggtttagg cgagaatgat 960
gagcaagtat atgcaacaat gaaagcactt cgtgaggcag atgtagactg cttgacttta 1020
ggacaatata tgacgccaac aaggcgtcac cttaaggttg aagaatatat tactcctgaa 1080
aaattcaaat actgggaaaa agtaggaaat gaacttggat ttcattatac tgcaagtggc 1140
cctttgtgct gttcttcata taaagcaggt gaatttttcc tgaaaaatct agtggctaaa 1200
agaaaaacaa aagacctcta aaacttcaac aagaccttca agatcacaga aatttttaaa 1260
atttgattcc agttaataac agaggtggtg ccagaatgcc tggactgcag tggatgtacc 1320
ccacctcttt gcttaaaaaa aaaaatgtca atagccaggc atagtggctc acgcctgtaa 1380
tcccagcact ttaggaggcc aaggcgggtg gatcacctga ggtcaggagt tcgagaccag 1440
cctggccaac atggtgaaat cctgtctcca ctaaaaacac aaaaattagt caggcgtggt 1500
agtgggtgcc tgtaatccca gctactcggg aggctaaggc aggagaatca cttgaacctg 1560
ggagggggag gttgcagtga gccaaagatc ctccattgcc ctccagcctg ggtgacaaga 1620
gca
1623

```

<210> 12

<211> 2199

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 72852842CB1

<400> 12

```

cagctatacc tcttttgaag attttaagaa cttagcctcc tgaacagtct tcttcgaaag 60
tgaaaagtgg taacagctga tgagtatcaa gaaattattt tctgcaaagg ggcagagtta 120
attgtatttg gaaccatga cagcacctac tggggaaaga cttctaagtg aggagaaacg 180
gctctacagg tcatgaaact atggaaatga gatggttttt gtcaaagatt caggatgact 240
tcagagggtg aaaaattaac ctagaaaaaa ctacagaggtt acttgaaaaa ttagatattc 300
ggtgcagtta tattcatgtg aaacagattt ttaaggacaa tgacaggctg aaacaaggaa 360
gaatcaccat agaagaattt agagcaattt atcgaattat cagcacacaga gaagaaatta 420
ttgagatttt caacacatat tctgaaaacc ggaaaattct tttagcaagt aatctggctc 480
aatttctgac acaagaacaa tatgcagctg agatgagtaa agctattgct tttgagatca 540
ttcagaaata cgagcctatc gaagaagtta ggaaagcaca ccaaagtgtc ttagaaggtt 600

```

```

ttacaagata catggattca cgtgaatgtc tactgtttta aaatgaatgt agaaaagtgt 660
atcaagatat gactcatcca ttaaatgatt attttatttc atcttcacat aacacatatt 720
tggtatctga tcaattattg ggaccaagtg acctttgggg atatgtaagt gcccttgtga 780
aaggatgccg ttgtttggag attgactgct gggatggagc acaaaatgaa cctgttgtat 840
atcatggcta cacactcaca agcaaaacttc tgtttaaaac tgttatccaa gctatacaca 900
agtatgcatt catgacatct gactaccagc tgggtgctctc tttagaaaat cactgctcca 960
ctgccccaca agaagtaatg gcagacaatt tgcaggctac ttttggagag tccttgcttt 1020
ctgatatgct tgatgatttt cctgatactc taccatcacc agaggcacta aaattcaaaa 1080
tattagttaa aaataagaaa ataggaaact taaaggaaac ccatgaaaga aaaggttctg 1140
ataagcgtgg taaggtggag gaatgggaag aagaagtggc agatggagag gaggaggagg 1200
aggaggagga ggaggaggag gaggaggagg aggataaatt caaagaatca gaagtattgg 1260
aatctgtttt aggagacaat caagacaagg aaacaggggt aaaaaagtta cctggagtaa 1320
tgcttttcaa gaaaaagaag accaggaagc taaaaattgc tctggcctta tctgatcttg 1380
tcattttatac gaaagctgag aaattcaaaa gctttcaaca ttcaagatta tatcagcaat 1440
ttaatgaaaa taattctatt ggggagacac aagcccgaaa actttcaaaa ttgcgagtcc 1500
atgagtttat ttttcacacc aggaagtcca ttaccagaat atatcccaa gcaacaagag 1560
cagactcttc taattttaat ccccaagaat tttggaatat aggttgtaa atggtggctt 1620
taaatttcca gacctctggt ctgccatgg atctgcaaaa tgggaaattt ttggataatg 1680
gtggttctgg atatattttg aaaccacatt tcttaagaga gagtaaata tactttaacc 1740
caagtaacat aaaagagggt atgccaatta cacttacaat aaggctcacc agtggtatcc 1800
agtgcctct tactcattca tcatctaaca aaggtgattc attagtaatt atagaagttt 1860
ttggtgttcc aatatgatac atgaagcagc agactcgtgt aattaaaaaa aatgctttta 1920
gtccaagatg gaatgaaaca ttcacattta ttattcatgt cccagaattg gcattgatac 1980
gttttgttgt tgaaggtcaa ggtttaatag caggaaatga atttcttggg caatatactt 2040
tgccacttct atgcatgaac aaaggttacc gtcgtattcc tctgttttcc agaatgggtg 2100
agagccttga gcctgcttca ctgtttgttt atgtttggta cgtcagataa cagctaata 2160
taaatgacat atcattagct atgcatcgca ataaaaccg 2199

```

&lt;210&gt; 13

&lt;211&gt; 6326

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7484271CB1

&lt;400&gt; 13

```

aggaagggga ggaggggggtg ggatattatg atgcagctca cattgaacac tctctttcct 60
gttgtttcca caccagctat tacgtatatt gtcaccgtct tcaactggga tgtccggggg 120
gctggtacca atatacgtaa tagctggtgt tgtaaccagg aaggggagga gggggtggga 180
tattatgatg cagctcacat tgaacactct ctttcctggt gtttcacac cagctattac 240
gtatatgtgc accgtcttca ctgggggatgt ccgggggggct ggtaccaa at ccaaaatcta 300
cttggtcatt tatggggcca gaggggataa gaacagtggg aaaatcttcc tggaggggcg 360
cgtgtttgac cgaggccgca cggacatctt ccacatcgag ctggctgtcc tccttagccc 420
cctgagtcgg gtctccgtcg ggcattggca tgtgggtgtc aacagaggct ggttctgtga 480
gaaggtgggt attctgtgcc ccttcactgg tatccagcag acctccctt gtagcaactg 540
gctggatgag aagaaagcgg atgggttgat cgagaggcag ctctatgaga tgggtgtctc 600
caggaagaag cggctgaaaa aattcccttg gtccctgtgg gtctggacaa ccgacctaaa 660
gaaagctggt accaactctc ccatcttcat ccagatttat gggcagaagg ggcggacaga 720
tgagattctc ctgaatccca acaacaagtg gttcaaacc ggcataatcg agaagtttag 780
gattgagctc ccggtacttg gcaaggttta taagattcga gtatggcatg ataaaaggag 840
ttctggttct ggatggcatt tagaaaggat gacctgatg aacactctga acaaagacaa 900
gtacaacttc aattgcaacc gctggctgga tgccaatgag gatgacaatg agatagttag 960
ggaaatgact gcagaaggcc caacagtgcg caggatcatg ggcattggccc ggtaccatgt 1020
gactgtgtgc acaggtgaac ttgaaggtgc tgggaccgat gccaacgtct atctctgcct 1080

```

ttttggtgat	gtgggggaca	cgggggaacg	gctgctctac	aactgcagga	ataacacaga	1140
cctgtttgaa	aagggcaatg	ctgacgagtt	cactatcgag	tctgtcacca	tgcggaatgt	1200
gagggcgggtg	aggatcagac	acgatggcaa	aggctccggc	agcggctggt	acctggacag	1260
agtgtctggtg	agagaggagg	ggcagcctga	gagcgacaac	gtggagttcc	catgtctcag	1320
gtggttggac	aaggataagg	atgatgggca	gctggtccga	gagttgctac	ccagtgcag	1380
cagcgcgaca	ctgaagaact	tctgctatca	catcagcttg	aagactgggg	atgtctctgg	1440
ggccagcacg	gattctagag	tctacatcaa	gctctatggg	gataaatctg	acaccatcaa	1500
gcaagttctt	cttgtctctg	acaacaacct	caaagactac	tttgaacgtg	gccgggtgga	1560
tgagttcacc	ctcgagaccc	tgaacattgg	aaatatcaac	cggctggtga	ttgggcatga	1620
cagcactggc	atgcatgcca	gctggttctt	gggcagcggt	cagatccgtg	tgccccgtca	1680
aggcaagcag	tacaccttct	ccgccaaccg	ctggctggac	aagaaccagg	ctgacgggcg	1740
cctggaggtg	gagctgtatc	ccagcgaggt	ggtggagatc	cagaaattgg	tccactatga	1800
ggttgagatt	tggacaggag	atgtgggtgg	cgcaggcacc	agtggcccag	tctacatgca	1860
gatctatgga	gagaaaggca	agacagaagt	gctcttcttc	tccagccgct	caaaagtttt	1920
tgaacgggcg	tccaaggaca	cattccagct	tgaggcggcc	gacgtgggcg	aggtctataa	1980
gctccggctc	gggcacacgg	gcgagggtt	tgggcccagc	tggttcgtgg	acaccgtgtg	2040
gctcgggcac	ctggtggtgc	gggaggtgga	cctcacgccc	gaggaggagg	cccggaagaa	2100
gaaggagaag	gacaagctgc	ggcagctgct	caagaaggag	cggctgaagg	ccaagctgca	2160
gaggaagaag	aagaagagga	agggcagcga	cgaagaggac	gagggggagg	aagaggagtc	2220
gtcctcatca	gaggagtctt	cgtcagagga	ggaggagatg	gaagaagagg	aggaagagga	2280
ggagtttggg	ccggggatgc	aggaggtgat	tgagcagcac	aagttcgaag	cccaccgctg	2340
gctggcccgg	ggcaaggagg	acaacgaact	tgtcgtggag	ttggtgccag	ctggcaagcc	2400
gggtcctgag	cgaaacacct	atgaggttca	ggtggtcacg	gggaatgtgc	ccaaggccgg	2460
cactgatgct	aacgtctacc	taaccatcta	cggcgaggag	tatggagaca	cgggcgaacg	2520
acccctgaag	aagtccagaca	agtccaacaa	atttgagcag	gggcagacag	acaccttcac	2580
catctatgcc	attgacctgg	gggcccctgac	caagattcgg	attcgccacg	acaacacagg	2640
caacagagca	ggttggttcc	tggacagaat	agacattact	gacatgaaca	acgagatcac	2700
gtactacttt	ccatgccaac	gttggtctggc	agtggaggaa	gatgatggcc	agctgtccag	2760
ggagctgttg	ccagtggatg	agtcttatgt	gctgccacag	agcgaggagg	gtgggggagg	2820
cgggtgacaac	aaccccctcg	acaacctggc	cctggagcag	aaagataaat	ctaccacatt	2880
ctcagtgacc	ataaagactg	gggttaagaa	gaatgcgggc	acagatgcta	atgtcttcat	2940
cacactcttt	ggcacacagg	atgacactgg	aatgaccctc	ctgaagtcct	ccaagacaaa	3000
cagcgataag	tttgagaggg	acagcattga	aatcttcacg	gtggagacgc	tggatctggg	3060
agacctgtgg	aaagtccggc	ttggccatga	caacacaggc	aaggccccag	gctgggttgt	3120
agactgggta	gaggtggatg	ccccatctct	tgggaagtgc	atgacgtttc	cctgtggccg	3180
ctggctggcc	aaaaacgaag	acgacgggtc	catcatcaga	gacctcttcc	atgcagagct	3240
tcagacgagg	ctgtacacac	catttgttcc	ttacgagatc	accctctaca	ccagtgtatg	3300
ctttgctgct	gggacagatg	ccaacatctt	catcatcatc	tatggctgcg	atgccgtgtg	3360
caccagcag	aagtatctgt	gtaccaacaa	gagggaaacg	aagcagttct	ttgagaggaa	3420
gtctgcctcc	cgcttcatcg	tagagttaga	agatgtggga	gaaatcattg	aaaaaattcg	3480
gattggccat	aataacacgg	gcatgaatcc	tgggtggcac	tgctctcacg	tggacatccg	3540
caggctcctc	ccggataaag	acggtgcaga	gaccttgact	ttcccatgcg	atcggtggct	3600
tgccacctct	gaggatgaca	aaaagacat	tcgagaactg	gttccatgat	acatcttcac	3660
tgagaaatac	atgaaagatg	ggtccttacg	gcaagtctac	aaggaagtag	aagagcctct	3720
ggacattgtg	ctgtactcgg	tgcagatctt	cacagggaac	attcctgggg	cagggacgga	3780
tgccaagggtg	tacatcacca	tctatggaga	cctcggggac	actggggagc	gataccttgg	3840
caagtcagag	aaccggacca	acaagttcga	gagagggaacg	gctgacacct	tcacatcaga	3900
ggcgcgtgac	ctaggcgtca	tctacaagat	caagctccgc	catgacaact	ccaagtgggtg	3960
cgcagactgg	tacgtggaga	aggtggagat	ctggaatgac	accaacgagg	acgagttcct	4020
gttcttatgc	gggcgctggc	tctccctgaa	gaaggaggat	gggcgactcg	agaggctctt	4080
ttacgagaag	gagtacactg	gggaccgcag	cagcaactgc	agcagccctg	ctgacttctg	4140
ggagatcgcc	ctgagctcca	agatggccga	tgtcgacatc	agcacagtga	ccgggcccac	4200
ggctgactac	gttcaagagg	gccaatttat	tcctactat	gtgtcagtca	ccactgggaa	4260
gcacaaggac	gcgccactg	acagccgagc	cttcatcttt	ctcatcgggg	aggatgatga	4320
acgtagtaag	cgcactctgt	tggactaccc	ccgagggaag	aggggcttca	gccgtggctc	4380
tgtggaggag	ttctacgtcg	caggcttggga	tgtgggcacg	atcaagaaaa	tagagctggg	4440

```

ccatgacggg gcctccctg agagctgctg gctggtggaa gagttgtgtt tggcagtgcc 4500
caccagggc accaagtaca tgttgaaactg taactgctgg ctggccaagg acagaggcga 4560
cggcattacc tcccgtgtct tgcacctctt ggatgccatg gtggtgaaca ttggggtgaa 4620
ggttctctat gaaatgacgg tgtggacagg ggatgtggtt ggcgggggca ctgactccaa 4680
catcttcatg accctctacg gcatcaacgg gagcacagag gagatgcagc tggacaaaaa 4740
gaaagccagg tttgagcggg agcagaacga caccttcatc atggagatcc tagacattgc 4800
tccattcacc aagatgcgga tccggattga tggcctgggc agtcggccgg agtggttcct 4860
ggagaggatc ctactgaaga acatgaacac tggagacctg accatgttct actatggaga 4920
ctggctgtcc cagcgggaagg gcaagaagac cctggtgtgt gaaatgtgtg ccgttatcga 4980
tgaggaagaa atgatggagt ggacctccta caccgtcgca gttaagacca gcgacatcct 5040
gggagcaggc actgatgcca acgtgttcat catcatcttc ggggagaacg gggatagtgg 5100
gacactggcc ctgaagcagt cggcaaacctg gaacaagttt gagcgggaaca acacggacac 5160
attcaacttc cctgacatgc tgagcttggg ccacctctgc aagctgaggg tctggcacga 5220
caacaaaggg atatttcctg gctggcatct gagctatgtc gatgtgaagg acaactcccg 5280
cgacgagacc ttccacttcc agtgtgactg ctggctctcc aagagtgagg gtgacgggca 5340
gacgggtccg gactttgcct gtgccaacaa caagatctgt gatgagctgg aagagaccac 5400
ctacgagatc gtcatagaaa cgggcaacgg aggcgaacac agggagaacg tctggctcat 5460
cctggagggc aggaagaacc gatccaaaga gtttctcatg gaaaattctt ctaggcagcg 5520
ggcctttagg aaggggacca cagacacgtt tgagtttgac agcatctact tgggggacat 5580
tgccctccctc tgtgtgggcc accttgccag ggaagaccgg tttatcccca agagagaact 5640
tgccctggcat gtcaagacca tcaccatcac cgagatggag tacggcaatg tgtacttctt 5700
taactgtgac tgcctcatcc ccctcaagag gaagaggaag tacttcaagg tattcgaggt 5760
taccaagacg acagagagct ttgccagcaa ggtccagagc ctggtgcccg tcaagtacga 5820
agtcacgtg acaacaggct atgagccagg ggcaggcact gatgccaacg tcttcgtgac 5880
catctttggg gccaacggag acacaggcaa gcgggagctg aagcagaaaa tgcgcaacct 5940
cttcgagcgg ggcagcacag accgcttctt cctggagacg ctggagctgg gtgagctgcg 6000
caagggtcgc ctggagcagc acagcagtgg ctactgctca ggctggctgg tggagaaggt 6060
ggaggtcacc aacaccagca cggcggtggc caccatcttc aactgtggca ggtggctgga 6120
caagaagcgg ggggatggac tcacctggag agacctcttc ccttctgtct gaggggctag 6180
ggccccacc ctctcactga gatgccccaa tctcacattt ctgccctcca ccttggcggt 6240
cagcagccct tcaaagcctc tagcattggc actgggggct agcagtcac tgagaacttc 6300
atggggtcct gctcccacc ctacc 6326

```

&lt;210&gt; 14

&lt;211&gt; 1561

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474074CB1

&lt;400&gt; 14

```

gctcgaccag tactcagctt gactgacatt ttcttatttc agataataaa agaccatgcc 60
ttgaattctc tcagctaagt gtaaaggatt ccttcagaga tttatttatt ccgagaatag 120
agaccattct gatgatgtat acaagggaaca acctaaactg tgctgagcca ctgtttgaac 180
aaaataactc acttaatggt aatttcaaca cacaaaagaa aacagtctgg cttattcacg 240
gatacagacc agtaggctcc atcccattat ggcttcagaa cttcgtaagg attttgctga 300
atgaagaaga tatgaatgta attgtagtag actggagccg ggggtgctaca acttttattt 360
ataatagagc agttaaaaac accagaaaag ttgctgtgag tttgagtgtg cacattaaaa 420
atcttttgaa gcatggtgca tctcttgaca attttcatct cataggtgtg agcttagggg 480
ctcatatcag tggatttggt ggaaagatat ttcattggtc acttgggaaga ataacaggtc 540
ttgacctgc tgggccaagg ttctccagaa aaccaccata tagcagatta gattacacgg 600
atgcaaagtt tgtggatgtc atccattctg actccaatgg tttaggcatt caagagccct 660
tgggacatat agatttttat ccaaattggag gaaataaaca acctggctgt cctaaatcaa 720
ttttctcagg aattcaattc attaaatgca accaccagag agcagttcac ttgttcatgg 780

```

```

catctttaga aacaaactgc aattttatgt catttccttg tcgttcatac aaagattaca 840
agactagctt atgtgtggac tgtgactgtt ttaaggaaaa atcatgtcct cggctgggtt 900
atcaagccaa gctattttaa ggtgttttaa aagaaaggat ggaaggaaga cctcttagga 960
ccactgtgtt tttggataca agtgggtacat atccattctg tacctattat tttgttctca 1020
gtataattgt tccagataaa actatgatgg atggctcgtt ttcattttaa ttattaaatc 1080
agcttgaaat gattgaagag ccaaggcttt atgaaaagaa caaaccattt tataaacttc 1140
aagaagtcaa gattcttgct caattttata atgactttgt aaatatttca agcattgggt 1200
tgacatattt ccagagctca aatctgcagt gttccacatg cacatacaag atccagagt 1260
tcatgttaaa atcacttaca tacccaaaaa gaccaccact ttgcaggtat aatattgtac 1320
ttaaagaaag agaggaagtg tttcttaatc caaacacatg tacgcaaaag aacacataag 1380
atgccttctt ccatcaaatg cacttgcttg tgaattaatg gacttgtaaa tgaaacaatg 1440
caatcagctt tttataatac actgttcaat ttgagattca agtatttcta tttcttgtaa 1500
aaaaatttaa gaatcaaaaa taaagaaaat aaaaagtgc aacagttaaa cattccaaaa 1560
a
1561

```

&lt;210&gt; 15

&lt;211&gt; 4941

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 72024970CB1

&lt;400&gt; 15

```

attccttggt ggccctggag ggtggatagg ctggcctggg ggccatcagg acagcaggtg 60
acggtcaggc caatgccagc cgggcctggg cacagccctg tgggggcttc ggagggccct 120
gaggaggagg aggaagaggc agaggagaga agggccacg gaggtcctgt cgccagcgct 180
gccactgcct gacctcgcct gcccgaaggc cgggtggcct ctgtggcctc cgtgaagcag 240
gcccggctgt cgtcaggcca tgtctggtcc atggccctcc cccgacagcc ggaccaaggg 300
aacgggtggc tggctggcgg aggtactcct ctgggtggag ggagtgtggt gctgtcttca 360
gagtggcagc tcggccccct ggtggagcgg tgcattgggt ccatgcaaga ggggatgcag 420
atggtgaagc tgcgtggcgg ctccaagggc ctggtccgct tctactacct ggacgagcac 480
cgctcctgca tccgctggag gccctcacgc aagaacgaga aggccaaagat ctccatcgac 540
tccatccagg aggtgagtga ggggcggcag tcggagggtc tccagcgcta ccctgacggc 600
agcttcgacc ccaactgctg ctccagcctc taccacggca gccaccgca gtcgctggac 660
ctggtctcca ccagcagcga ggtggcgcgc acctgggtca ctggcctgca ctacctcatg 720
gccggcatca gcgacgagga cagcctggct cgcccgagc gcaccaggga ccagtggctg 780
aagcagacgt ttgacgaggc cgacaagaac ggggatggca gcctgagcat tggcgaggtc 840
ctgcagctgc tgcacaagct caacgtgaac ctgccccggc agaggggtgaa gcagatgttc 900
aggggaagcgg acacggatga ccaccaaggg acgctgggtt ttgaagagtt ctgtgccttc 960
tacaagatga tgtccacccg ccgggacctc tacctgctca tgctgacctc cagcaaccac 1020
aaggaccacc tggatgccgc cagcctgcag cgcttcctgc aggtggagca gaagatggcg 1080
ggtgtgacct tcgagagctg ccaggacatc atcgagcagt ttgagccatg cccagaaaac 1140
aagagtaagg ggctgctggg cattgatggc ttcaccaact acaccaggag ccctgctggg 1200
gacatcttca accctgagca ccacatgtg caccaggaca tgacgcagcc gctgagccac 1260
tacttcatca cctcgtccca caacacctac ctggtgggtg accagctcat gtcccagtca 1320
cggggtggaca tgtatgcttg ggtcctgcag gctggctgcc gctgcgtgga ggtggactgc 1380
tgggatgggc ccgacgggga gccattgtg caccatggct acactctgac ttccaagatc 1440
ctcttcaaag acgtcattga aaccatcaac aaatatgcct tcatcaagaa tgagtacca 1500
gtgatcctgt ccatcgaaaa cactgcagc gtcacccagc agaagaaaat ggcccagtat 1560
ctgactgaca tccttgggga caagctggac ctgtcatcag tgagcagtgat agatgccacc 1620
acactccctt ctccacagat gctcaagggc aagatcctcg tgaaggggaa gaagctccca 1680
gccaacatca gcgaggatgc ggaggaaggc gaggtgtctg atgaggacag tgctgatgag 1740
attgacgatg actgcaagct cctcaatggg gatgcaccca ccaatcgaaa gcgtgtagaa 1800
aacactgcta agaggaaact ggattccctc atcaaagagt cgaagattcg ggactgtgag 1860

```

```

gaccccaaca acttctccgt ctccacactg tccccatctg gaaagctcgg acgcaagagc 1920
aaggctgaag aggacgtgga gtctggggag gatgccgggg ccagcagacg caatggccgc 1980
ctcgtcgtgg gaagcttctc caggcgcaag aagaagggca gcaagctgaa gaaggcgcc 2040
agcgtggagg agggagatga gggtcaggac tccccgggag gccagagccg aggggagacc 2100
cggcagaaga agaccatgaa gctgtcccg gcccctctctg acctggtgaa gtacaccaag 2160
tccgtggcca cccacgacat agagatggag gcggcgtcca gctggcaggt gtcgtccttc 2220
agcgagacca agggccacca gattctgcag cagaagccgg cgcagtacct acgcttcaac 2280
cagcagcagc tctcccgcat ctacccctcc tcctaccgtg tggactccag caactacaac 2340
ccgcagccct tctggaacgc cggctgcca atggttgccc tgaactacca gtcagagggg 2400
cggatgctgc agctgaaccg agccaagttc agcggcaacg gtggctgcgg ctacgtactc 2460
aagcctgggt gcatgtgcca gggcgtgttc aaccccaact cggaggaccc cctgcccggg 2520
cagctcaaga agcagctggg gctccggatc atcagtggcc agcagcttcc caagccgcgc 2580
gactccatgc tgggggaccg tggggagatc atcgacccct ttgtggaggt ggagatcatt 2640
gggctccctg tggactgcag caggagcag acccgctgg tggacgacaa cgggttcaac 2700
cccacctggg aggagaccct ggttttcatg gtgcacatgc cggagatcgc gctggctcgc 2760
ttctcgtctc gggaccacga tcccccggg cgtgacttca ttggccagag gacgtggcc 2820
ttcagcagca tgatgccagg ctacagacac gtgtacctag aagggatgga agaggcctcc 2880
atcttcgtgc atgtggctgt cagtacatc agcggtaagg tcaagcaggc tctgggccta 2940
aaaggcctct tctcccgagg cccaaagccc ggctcgtgg acagtcatgc tgctgggcgg 3000
cccccgccc gggcctccgt tagccagcgg atcctgcggc gcacggccag cgccccgacc 3060
aagagccaga agccgggccc caggggcttc ccggagctgg tccctgggtac acgggacaca 3120
ggctccaagg ggggtggcaga cgaatgtggt cccccgggg ccggacctgc tccggaagcc 3180
ccagcccagg agggggcccgg cagcggcagc ccccgaggta aggcgccagc tgcggtggca 3240
gagaagagcc ctgtgcgagt gcggcccccg cgtgtcctgg acggcccccg gcctgctggg 3300
atggccgcca atgcatgaa gtgtgtggtg ggaatcctgc ccggcgtgaa caccgggggc 3360
ccgagagggg agggccacc cagcccgggg cctgcaagca ggcaggcagc cattcgccag 3420
cagccccggg cccgggctga ctactgggg gcccctgct gtggcctgga ccctcacgct 3480
atccccggga gaagcagaga ggcccccaag ggtcctgggg cctggaggca gggctccaggc 3540
ggtagcggct ccatgtctc ggactccagc agcccagaca gcccgggcat ccccgaaagg 3600
tcccccgct ggcctgaggg tgcctgcagg caaccggggg ccctgcaggg agagatgagt 3660
gccttggttg ctcaaaagct ggaggagatc aggaagaaat ccccatggt tcccgccgtt 3720
aggaactgag agcggcgagt gacagacacc cgccccctct ccacgcagcg gccactcccc 3780
ccactgtgca gcctggaac catcgtgag gagcccgccc caggccctgg tcccccgcca 3840
ccagcggctg tccccaccag ctcttctcag ggacggcccc cataccccac aggaccggga 3900
gccaatgtgg caagccccct agaggacact gaggagcccc gagacagcag gcctcgccg 3960
tgcaacggcg agggcgccg cggggcatac gagagggccc ccggcagcca gacggacggc 4020
aggagccagc cccggaccct gggccacctg cccgtgatta gaagggtgaa gaggtagggg 4080
caggtgccca cggagccctt gggaggggtg cggccccctg ccgctccctt tccagctcct 4140
gccgtgtact ccgatgccac gggcagtgac ccgctgtggc agcggctgga gccatgtggc 4200
caccgagaca gcgtttctc ctctccagc atgtcatcca gcgacactgt cattgacctc 4260
tccctgcca gcctgggctt gggccgcagc cgtgagaacc tcgctggagc ccacatggga 4320
cgctgcccc ccaggcccca ctcggtctc gctgcccgc cagacctgcc acctgtgacc 4380
aagagcaaat ccaaccccaa ccttcgggct acaggccagc ggcctcccat acctgacgaa 4440
ctgcagccca ggtccctggc cccaaggatg gctggcctcc ccttcgggc tccctggggc 4500
tgcccttccc tgggtggcgt gcaggactgc cccgtggctg ccaagtccaa gacctgggc 4560
gacctcactg ctgatgactt tgcccctagc tttgagggcg gctcccgcag actgagccac 4620
agcctggggc tcccgggagg gacacggcgg gtgtcggggc cagggtgaga cgggacaccc 4680
tgacagagca gctgcgctgg ctactgtct tccagcaggc aggagacatc acgtcaccca 4740
ccagcctggg cccggctggg gagggggtgg caggggcccc ggttttgtgc ggcgctcctc 4800
ctcccgagc acagcgctg cgtgcattgc agcggccgca ggccggagcg cgcgaactga 4860
ggctggggcg gggaccagag gggggggccc cggggccgtt ctggccaggt gtgtttccgc 4920
tcagggtcgg ctctttacc t
4941

```

&lt;210&gt; 16

&lt;211&gt; 4159

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6131380CB1

&lt;400&gt; 16

tctcgggtctt	cggtgcgaga	tcactttggt	cctggagaca	gtgtacagag	cttggaattc	60
tctgaggggt	tcctgagatg	gcgccattca	agccaggggg	ttgaatagct	tgactcttca	120
tttcagcagc	acgattgacc	cctcagtggg	tcagcagcga	ttcattccac	acgtatttag	180
ggtccttggt	gaattttgtc	atggttattt	aaggaacctt	gcctagaagt	cccaacttgc	240
agttcccat	cgacgggaag	gcttggactc	caagatgatt	ataaaggaat	atcggattcc	300
tctgccaatg	accgtggagg	agtaccgcat	cgcccagctg	tacatgatac	agaagaagag	360
ccgtaacgag	acatatggcg	aaggcagcgg	cgtggagatc	ctggagaacc	ggccgtacac	420
agatggccca	ggcggctctg	ggcagtagac	acacaagggt	tatcatgtgg	gcatgcacat	480
tcccagctgg	ttccgctcca	tcctgcccac	ggcagccctg	cggttggtgg	aggagtcttg	540
gaatgcctac	ccctacaccc	gaaccagggt	cacctgtcct	ttcgtggaga	aattctccat	600
cgacattgaa	aaccttttata	aaactgatgc	tggagaaaac	cccagactgt	tcaacctctc	660
tcctgtggaa	aagaaccagc	tgacaatcga	cttcacagac	attgtcaaag	acctctgtgc	720
ccacaacgag	tataagacag	aagaggaccc	caagctgttc	cagtcaacca	agaccacagc	780
ggggcccttg	tccgagaact	ggatcgagga	gtacaagaag	caggtcttcc	ccatcatgtg	840
cgcatacaag	ctctgcaagg	tggagtcccg	ctactggggc	atgcagtcca	agatcgagag	900
gttcatccac	gacaccggac	tacggagggt	gatggtgcgg	gctcaccggc	aggcctggtg	960
ctggcaggac	gagtgggtatg	ggctgagcat	ggagaacatc	cgggagctgg	agaaggaggc	1020
acagctcatg	ctttcccgtg	agatggccca	gttcaatgag	gatggtgagg	aggccactga	1080
gctcgtcaag	cacgaagccg	tctcggacca	gacctctggg	gagcccccg	agcccagcag	1140
cagcaatggg	gagcccctag	tggggcgcg	cctcaagaaa	cagtgttcca	catcctccaa	1200
gtcgtctcgg	tcgtccaagc	ggggagcag	tccttcccgc	cacagcatct	cagagtggag	1260
gatgcagagt	attgccagg	actcggatga	gagctcagat	gatgagttct	tcgatggcca	1320
cgaggacctg	tccgacacag	aggaaatgtt	ccccaaagg	atcaccaagt	ggagctccaa	1380
tgacctcatg	gacaagatcg	agagcccaga	gccggaagac	acacaagatg	gtctgtaccg	1440
ccagggtgcc	cctgagttca	gggtggcctc	cagtgtggag	cagctgaaca	tcatagagga	1500
cgagggttagc	cagccgctgg	ctgcaccgcc	ctccaagatc	cacgtgctgc	tactggtgct	1560
gcacggaggc	accatcctgg	acacaggcgc	cggggacccc	agctccaaga	agggcgatgc	1620
taacaccatc	gccaacgtgt	tcgacaccgt	catgcgcgtg	cactacccca	gcgccttggg	1680
ccgecttgcc	atccgcctgg	tgccctgccc	gcccgtctgc	tctgacgcct	ttgccctggg	1740
ctccaacctc	agccccata	gccatgacca	agctgtctgc	tccagcagtc	aggaccacat	1800
tccccctgct	gccctcccc	tgctggccac	ctcctcccc	cagtaccagg	aggcagttgc	1860
cacagtgatt	cagcgagcca	accttgccct	tggggacttc	atcaagtccc	aggagggcat	1920
gaccttcaat	gggcagggtc	gcctgattgg	ggactgcgtc	gggggcatcc	tggcatttga	1980
tgccctgtgc	tacagtaacc	agccgggtgc	tgagagtcag	agcagcagcc	gccggggcag	2040
cgtggtcagc	atgcaggaca	atgacctgct	gtccccgggc	atcctgatga	atgcagcaca	2100
ctgctgcggg	ggtggcggtg	gcggcggtgg	cggtggtggc	agcagtgggtg	gtggtggcag	2160
tagtggtggc	tccagcctgg	agagcagtcg	gcacctgagc	cgaagcaacg	tcgacatccc	2220
ccgcagcaac	ggcactgagg	accccaaaa	gcaactgccc	cgcaagagga	gcgactcatc	2280
cacctacgag	ctggatacca	tccagcagca	ccaggccttc	ctgtccagcc	tccatgccag	2340
cgtgctgagg	actgagccct	gtcacgccca	ttccagcagc	tccaccatgc	tggatggcac	2400
aggtgccctg	ggcaggtttg	actttgagat	caccgacctc	ttcctcttcg	ggtgcccgtc	2460
ggggctgggtc	ctggccctga	ggaagactgt	catcccagcc	ctggatgttt	tccagctgcg	2520
gccggcctgc	cagcaagtct	acaacctctt	ccaccccgcg	gaccctgcag	cttcacgcct	2580
ggagccgctg	ctggaacggc	gctttcacgc	cctgcccctc	ttcagcgtcc	cccgtacca	2640
acgctacccg	ctgggggatg	gctgctccac	gctgctggat	gtgctccaga	cccacaatgc	2700
agccttccaa	gagcatggcg	ccccctcctc	gccgggcact	gcccctgcc	gtcgtggctt	2760
ccgccgagcc	agtggatca	gcacgcgag	ccagggtgca	ggcatggctg	agagctacac	2820
ggcatccagc	atcgcccagg	tcgctgcaaa	gtggtggggc	cagaagcgga	tcgactacgc	2880
cctgtactgc	cctgacgccc	tcacggcctt	ccccacggtg	gctctgcctc	acctcttcca	2940



```

cgccagctac tgggagctcaa cagacgtggt ctcctttctg ctgagacagg tcatgaggca 3000
tgacaactcc agcatcttgg agctggatgg caaggaagtg tcggtgttca cccctcaaa 3060
gccaagggag aagtggcagc gcaagcggac ccacgtgaag ctgcggaacg tgacggccaa 3120
ccaccggatc aatgatgccc ttgccaatga ggacggcccc caggttctga cgggcagggt 3180
catgtatggg cccctggaca tggtcaccct gactggggag aaggtggatg tgcacatcat 3240
gaccagccg cctcaggcg agtggctcta cctggatacg ctggtgacca acaacagtgg 3300
gcgtgtctcc tacaccatcc ctgagtcgca ccgcctgggc gtgggtgtct accctatcaa 3360
gatgggtggtc aggggagacc acacgtttgc cgacagctac atcacctgc tgcccaaggg 3420
cacagagttc gtggtcttca gcatcgacgg ttcctttgcc gctagcgtgt ccatcatggg 3480
cagcgacccc aaggtgcggg ccggggccgt ggacgtggtg cggcactggc aggacctggg 3540
ctacctcatc atctacgtga cgggccggcc cgacatgcag aagcagcggg tgggtggcgtg 3600
gctggcccag cacaacttcc cccatggcgt ggtgtccttc tgtgacggcc tgggtgcatga 3660
cccgctgcgg cacaaggcca acttctgaa gctgctcatc tccgagctgc acctgcgcgt 3720
gcacgcggcc tatggctcca ccaaggacgt ggcggtgtac agcgccatta gcctgtcccc 3780
catgcagatc tacatcgtgg gccggccac caagaagctg cagcagcagt gccagttcat 3840
cacggatggc tacgcggccc acctggcgca gctgaagtac agccaccggg cgcggcccg 3900
tcgcaacacg gccaccgcga tggcgctgcg caagggcagc ttcggcctgc ccggccaggg 3960
cgactttctg cgctcccgga accacctgct tcgcaccatc tcggccacgc ccagcgggcc 4020
cagccaccgg cagcagcgga cacagagcca ggcggatggc gagcagcggg gccagcgag 4080
catgagtgtg gcggccggct gctggggccg cgccatgact ggccgcctgg agccgggggc 4140
agccgcgggc cccaagtag
4159

```

&lt;210&gt; 17

&lt;211&gt; 1481

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 643681CB1

&lt;400&gt; 17

```

atttgtgttaa tgtctttgtc tccattagac ctttattatt tgatttacgt ctggctctga 60
actcctgacc tcaggtggtc caccgcctc ggctcccaa agtgctggca ttacaggcgt 120
gaaccaccgt gcctggccgg aagtctttaa aaaataaagt gattctactc ttctaagctt 180
acagagacca gaccagtgga atgtaactgg gaaaaatcaa gatggtacct ctctgcatta 240
tcccgcgaga cactgtattt tatgcattca tgtctaggat acagtgtgaa aattaaag 300
tttagagggc agatgcaatt gtggcaagt acctgccaat aaagcagggt cagctataga 360
agctggcata ggtatattcct taatggtgct ttctccctgg gcttgtcttt ttgtgtttt 420
ttttccctta tattcagaag ctcttgaga agtgataaac acctccagct ttctaacatc 480
ctccccacac catctacca tatccatctc ccagcatcca tctgcattca gctaagggcg 540
ggaaactgac ctagtgcctg tgttgacagc catttctgag gtctccacca tccaaggagg 600
cacagccgtc attactgtcc tccatgcctt cagcagcccc cctcacagct aaggtacata 660
ccacccttc tgccgcgcct ccaccctgg caccaaggtc ttctgctgct tatgtctaaa 720
gggatcacct atatttaact gcctcagtga cctaacctct ttcttctcat gtgccagatg 780
ttaagatgaa ggaggaatac aacacatact caagcctcag cctgtttagt tgttttact 840
ggggctcgct ttctgggac ggtatttatt atcagactgg caagcctaac tccataggtt 900
tacaggaagt agggatattt ttataaaca attgtgctc cccacattt tgctatgtta 960
atatttgctt ctaacaattt gcagctgttt cactttttcc tcatttgtct ctaagtgtga 1020
ggctttgttg gaggggacag agcacaggaa cagccttgac agtctgtaat tattgtacag 1080
atattttaat agcatataaa taagtatatt ccttttattt tgaaacaaaa atgatcagac 1140
actgcctttt gtgtgtttgc tgctgtggc atcctttttt aaaaagactg ttacatatta 1200
aaatagtgtg catatataaa tattacctct tttgctgtac agttgtgata gagactgaag 1260
attttatttt ttgtgtgctt ttataagaa aaaaattaat acactaaaga atcttgcgtg 1320
tgtgattgtg atgtacctat gtaacttatt tacttttgaa tgttcttctg tatctttaa 1380
ccttttatta aataagggtt taaaaattaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1440

```

aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa c

1481

&lt;210&gt; 18

&lt;211&gt; 1841

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6897474CB1

&lt;400&gt; 18

```
gctctgctca gttctctgtg cctgtctccc tccagcactg ccgaggttct ctgccgaggc 60
caaccagaaa tacccttggg aagctggaat cctgcaacaa tggcccaggg tgctctctgg 120
atcctactcg gattgtact gtgtcagac ccagggacag cctccctgcc cctgctcatg 180
gactctgtca tccaggccct ggctgagctg gaggagaaag tgccagctgc caagaccaga 240
cacacagctt ctgctgtggc gatgtcagct ccaaactctg gccccacaa tcgcctctac 300
cacttctctg tgggggcatg gagcctcaat gctacagagt tggatccctg cccactaagc 360
ccagagctgt taggcctgac caaggaggtg gcccgacatg acgtacgaga agggagaggaa 420
tatgggggtg tgctggcacc tgatggctcg accgtggctg tggagcctct gctggcgggg 480
ctggaggcag ggctgcaagg gcgcaggggc ataaatttgc ccttgagacg catggctgcc 540
ccttgggaga ctggagatac ctttccagat gttgtggcca ttgctccaga tgtaagagcc 600
acctcctccc caggactcag ggatggctct ccagatgtca ccactgcaga tattggagcc 660
aacactccag atgtacaaa aggtgttcca gatgtccaag cttccttgcc agatgcaaaa 720
gccaagtccc caccgacat ggtggacagc ctcttgccag tcacctggc tggaaacctg 780
ggcctgacct tcctccgagg ttccagacc cagagccatc cagacctggg aactgagggc 840
tgctgggacc agctctctgc ccctcgacc ttacgcttt tggaccccaa ggcattctct 900
ttaaccatgg cttcctcaa tggcgccctg gatgggttca tccttgaga ctactgagc 960
cggactcctg agccccggcc atccctcagc cacttgctga gccagtacta tggggctggg 1020
gtggccagag acccaggggt ccgcagcaac ttccgacggc agaacgggtg tgctctgact 1080
tcagcctcca tcctggccca gcaggtgtgg ggaacccttg tccttctaca gaggtggag 1140
ccagtacacc tccagcttca gtgcatgagc caagaacagc tggcccaggt ggctgccaat 1200
gctaccaagg aattcactga ggccttctct ggatgcccg ccatccacce ccgtgccgc 1260
tggggagcgg cgcttatcg gggccgccc aagctgtctg agctgccgt gggattcttg 1320
tacgtgcac acacctagt gcctgcacca ccctgcacgg acttcaagcg ctgcgcagcc 1380
aacatgcgct ccatgcagcg ctaccaccag gacacgcaag gctggggaga catcggctac 1440
agtttcgtgg tgggctcgga cggctacgtg tacgagggac gcggctggca ctgggtgggc 1500
gcccacagc tcggccacaa ctcccggggc ttccggcgtg ccatagtggg caactacacc 1560
gcggcgctgc ccaccgagc cgctctgcgc acggtgcgag acacgctccc gagttgtgcg 1620
gtgcgcgccc gcctcctgcg gccagactac gcgctgctgg gccaccgcca gctgggtgcg 1680
accgactgcc ccggcgagc gctcttcgac ctgctgcgca cctggccgca cttcaccgcg 1740
actgttaagc caagacctgc caggagtgtc tctaagagat ccaggaggga gccaccccca 1800
aggacctgc cagccacaga cctccaataa agacagcatg g 1841
```